```
OM nucleic - nucleic search, using sw model

Run on:          October 14, 2003, 15:48:26 ; Search time 23109 Seconds
                                              (without alignments)
                                              10933.318 Million cell updates/sec

Title:           US-09-754-997A-1
Perfect score:   6176
Sequence:        1 atggcgcgggcggacacggg..........taataaactattttccagac 6176

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6176	100.0	6176	6	AX191259	AX191259 Sequence
	2	6176	100.0	6176	10	AF176694	AF176694 Mus muscu
	3	6038.8	97.8	6219	10	AB052620	AB052620 Mus muscu
	4	6025.8	97.6	6222	10	AB052621	AB052621 Mus muscu
	5	5810.2	94.1	6301	10	AK122535	AK122535 Mus muscu
	6	3756	60.8	3756	6	AX191303	AX191303 Sequence
	7	3396.6	55.0	6485	9	AB052622	AB052622 Homo sapi
	8	2796	45.3	2796	6	AX191261	AX191261 Sequence
c	9	2743.2	44.4	115135	10	AC112161	AC112161 Mus Muscu
c	10	2743.2	44.4	158055	2	AC110235	AC110235 Mus muscu
	11	2691	43.6	5549	9	AB046848	AB046848 Homo sapi
	12	2681.8	43.4	3741	6	AX418406	AX418406 Sequence
c	13	1843.2	29.8	224227	2	AC106509	AC106509 Rattus no
c	14	854.8	13.8	176282	2	AC068507	AC068507 Homo sapi
	15	854.8	13.8	198295	9	AC105129	AC105129 Homo sapi
	16	825	13.4	825	6	AX191263	AX191263 Sequence
c	17	659.4	10.7	72149	2	AC122776	AC122776 Mus muscu
	18	632.2	10.2	2560	9	BC040648	BC040648 Homo sapi
	19	571.2	9.2	72149	2	AC122776	AC122776 Mus muscu
c	20	415	6.7	116580	10	AC112162	AC112162 Mus Muscu
	21	378.8	6.1	216302	2	AC107847	AC107847 Mus muscu
	22	317	5.1	2967	9	BC042054	BC042054 Homo sapi
	23	316.6	5.1	2976	6	AR237558	AR237558 Sequence
	24	316.6	5.1	3210	6	AR237553	AR237553 Sequence
	25	316.6	5.1	3219	6	AR237561	AR237561 Sequence
	26	316.6	5.1	3874	6	AR237568	AR237568 Sequence
	27	316.4	5.1	2700	6	AX497857	AX497857 Sequence
	28	315.4	5.1	3453	6	AR237556	AR237556 Sequence
	29	308.4	5.0	3198	10	BC053057	BC053057 Mus muscu
	30	306.8	5.0	3146	10	AF026465	AF026465 Mus muscu
	31	288	4.7	288	6	AX191277	AX191277 Sequence
	32	287.4	4.7	2729	9	AK095529	AK095529 Homo sapi
	33	252	4.1	252	6	AX191281	AX191281 Sequence
	34	249	4.0	249	6	AX191273	AX191273 Sequence
	35	249	4.0	249	6	AX191275	AX191275 Sequence
	36	246	4.0	246	6	AX191279	AX191279 Sequence
	37	243	3.9	243	6	AX191265	AX191265 Sequence
	38	217	3.5	2481	6	AR237560	AR237560 Sequence
	39	217	3.5	2715	6	AR237555	AR237555 Sequence
	40	217	3.5	2724	6	AR237562	AR237562 Sequence
	41	217	3.5	2958	6	AR237557	AR237557 Sequence
c	42	215.8	3.5	147760	9	AC011846	AC011846 Homo sapi
	43	215.8	3.5	176282	2	AC068507	AC068507 Homo sapi
	44	213.8	3.5	2139	6	AR237563	AR237563 Sequence
	45	213.8	3.5	2382	6	AR237566	AR237566 Sequence

ALIGNMENTS

```
RESULT 1
AX191259
LOCUS       AX191259                6176 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 1 from Patent WO0149714.
ACCESSION   AX191259
VERSION     AX191259.1  GI:15209510
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 1 12-JUL-2001;
```

```
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..6176
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             1..3759
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51120.1"
                     /db_xref="GI:15209511"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

  Query Match            100.0%;  Score 6176;  DB 6;  Length 6176;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 6176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy     1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy     1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy     1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy     1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy     1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy     1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy     1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy     2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy     2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy     2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy     2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy     2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy     2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy     2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy     2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy     2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy     2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy     2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy     2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy     2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy     2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy     2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy     2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy     3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy     3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy     3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy     3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy     3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy     3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy     3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy     3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy     3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy     3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy     3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy     3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy     3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
```

```
Qy      3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840

Qy      3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900

Qy      3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960

Qy      3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020

Qy      4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080

Qy      4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140

Qy      4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200

Qy      4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260

Qy      4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320

Qy      4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380

Qy      4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440

Qy      4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500

Qy      4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560

Qy      4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620

Qy      4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680

Qy      4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740

Qy      4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800

Qy      4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860

Qy      4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920

Qy      4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980

Qy      4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040

Qy      5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100

Qy      5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160

Qy      5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220

Qy      5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280

Qy      5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340

Qy      5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400

Qy      5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460

Qy      5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520

Qy      5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580

Qy      5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640

Qy      5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700

Qy      5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760

Qy      5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820

Qy      5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880

Qy      5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940

Qy      5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000

Qy       6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060

Qy       6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120

Qy       6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176

RESULT 2
AF176694
LOCUS       AF176694                6176 bp    mRNA    linear   ROD 27-APR-2000
DEFINITION  Mus musculus neighbor of Punc e11 protein (Nope) mRNA, complete
            cds.
ACCESSION   AF176694
VERSION     AF176694.1  GI:7650185
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6176)
  AUTHORS   Salbaum,J.M. and Kappen,C.
  TITLE     Cloning and expression of nope, a new mouse gene of the
            immunoglobulin superfamily related to guidance receptors
  JOURNAL   Genomics 64 (1), 15-23 (2000)
  MEDLINE   20175427
   PUBMED   10708514
REFERENCE   2  (bases 1 to 6176)
  AUTHORS   Salbaum,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) The Neurosciences Institute, 10640 John Jay
            Hopkins Drive, San Diego, CA 92121, USA
FEATURES             Location/Qualifiers
     source          1..6176
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="band D-E1; 24.59 cR from D9Mit143"
     gene            1..6176
                     /gene="Nope"
     CDS             1..3759
                     /gene="Nope"
                     /note="immunoglobulin superfamily; related to guidance
                     receptors"
                     /codon_start=1
                     /product="neighbor of Punc e11 protein"
                     /protein_id="AAF65930.1"
                     /db_xref="GI:7650186"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
     misc_feature    1..23
                     /gene="Nope"
                     /note="unable to sequence region in mRNA"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

  Query Match             100.0%;  Score 6176;  DB 10;  Length 6176;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
```

```
Qy   3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy   3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy   3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy   3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy   3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy   3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy   3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy   3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy   3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy   3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy   3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy   3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780

Qy   3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840

Qy   3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900

Qy   3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960

Qy   3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020

Qy   4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080

Qy   4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140

Qy   4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200

Qy   4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260

Qy   4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320

Qy   4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380

Qy   4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440

Qy   4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500

Qy   4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560

Qy   4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620

Qy   4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680

Qy   4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740

Qy   4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800

Qy   4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860

Qy   4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920

Qy   4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980

Qy   4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040

Qy   5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100

Qy   5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160

Qy   5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220

Qy   5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTA 5280

Qy        5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340

Qy        5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400

Qy        5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460

Qy        5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520

Qy        5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580

Qy        5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640

Qy        5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700

Qy        5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760

Qy        5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820

Qy        5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880

Qy        5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940

Qy        5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000

Qy        6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060

Qy        6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120

Qy        6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176


RESULT 3
AB052620
LOCUS       AB052620                6219 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36 mRNA, complete cds.
ACCESSION   AB052620
VERSION     AB052620.1  GI:11862938
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6219)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6219
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
                     /clone="DDM36"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)~vector:pDDM36ORF"
     gene            1..6219
                     /gene="DDM36"
     CDS             50..3808
                     /gene="DDM36"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36"
                     /protein_id="BAB19278.1"
                     /db_xref="GI:11862939"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1319 a   1772 c   1652 g   1476 t
ORIGIN

  Query Match             97.8%;  Score 6038.8;  DB 10;  Length 6219;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 6136;  Conservative    0;  Mismatches   32;  Indels    8;  Gaps    7;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 109

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 169
```

```
Qy    121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    170 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 229

Qy    181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    230 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 289

Qy    241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 349

Qy    301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 409

Qy    361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    410 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 469

Qy    421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 529

Qy    481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 589

Qy    541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
          ||||||||||||||| ||||||||||||||||||||||  |  |||||||||||||||||
Db    590 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 649

Qy    601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 709

Qy    661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 769

Qy    721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    770 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTA 829

Qy    781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 889

Qy    841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 949

Qy    901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| 
Db    950 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGAT 1009

Qy    961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1010 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCC 1069

Qy   1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1129

Qy   1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1189

Qy   1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1190 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGC 1249

Qy   1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1309

Qy   1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1369

Qy   1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1429

Qy   1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1489

Qy   1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1549

Qy   1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1609

Qy   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1669

Qy   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1729

Qy   1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1730 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1789

Qy   1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1849

Qy   1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1850 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1909

Qy   1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1969

Qy   1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 2029

Qy   1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   2030 AAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2089

Qy   2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2090 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2149

Qy   2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2150 TATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTCAAC 2209

Qy   2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2210 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2269

Qy   2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2270 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2329

Qy   2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
```

```
Db  2330 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2389
Qy  2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
Db  2390 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2449
Qy  2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
Db  2450 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2509
Qy  2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
Db  2510 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2569
Qy  2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
Db  2570 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2629
Qy  2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
Db  2630 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2689
Qy  2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
Db  2690 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2749
Qy  2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
Db  2750 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2809
Qy  2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
Db  2810 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2869
Qy  2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
Db  2870 ACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2929
Qy  2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
Db  2930 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACGAAGC 2989
Qy  2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
Db  2990 TCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3049
Qy  3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
Db  3050 TACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3109
Qy  3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
Db  3110 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3169
Qy  3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
Db  3170 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3229
Qy  3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
Db  3230 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3289
Qy  3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
Db  3290 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3349
Qy  3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
Db  3350 CAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAGAAG 3409
Qy  3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
Db  3410 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3469
Qy  3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
Db  3470 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3529
Qy  3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
Db  3530 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3589
Qy  3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
Db  3590 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3649
Qy  3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
Db  3650 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3709
Qy  3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
Db  3710 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3769
Qy  3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
Db  3770 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3829
Qy  3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840
Db  3830 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3889
Qy  3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900
Db  3890 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3949
Qy  3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960
Db  3950 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 4009
Qy  3961 GTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
Db  4010 GTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4069
Qy  4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080
Db  4070 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4129
Qy  4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140
Db  4130 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4189
Qy  4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200
Db  4190 AAAGACAGCGCCAGCCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4249
Qy  4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260
Db  4250 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4309
Qy  4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320
Db  4310 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4369
Qy  4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380
Db  4370 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4429
Qy  4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440
Db  4430 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4489
Qy  4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500
Db  4490 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4549
```

```
Qy       4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4550 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4609

Qy       4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4610 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4669

Qy       4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||| |||||
Db       4670 CTGCCAACCGCCTTCCTGGGCAGATGA-CCCACTCCAAGTAGATCTGGAAAAGA-CCTTG 4727

Qy       4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4728 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4787

Qy       4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4788 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTG- 4846

Qy       4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4847 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4906

Qy       4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4907 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4966

Qy       4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4967 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 5026

Qy       4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5027 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5086

Qy       5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5087 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5146

Qy       5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5147 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5206

Qy       5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5207 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5266

Qy       5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5267 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5326

Qy       5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5327 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5386

Qy       5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5387 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5446

Qy       5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5447 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5506

Qy       5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5507 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5566

Qy       5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       5567 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTTTTTTCCTCC 5626

Qy       5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5627 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5686

Qy       5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700
              |||||||||||||||||||||||||||    | |   |||||||||||||||||||||||
Db       5687 TCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAA 5744

Qy       5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTA 5759
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       5745 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTA 5804

Qy       5760 ATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGC 5819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5805 ATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGC 5864

Qy       5820 ATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCAC 5879
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       5865 ATTATATATGATTTGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCAG 5924

Qy       5880 CAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGC 5939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5925 CAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGC 5984

Qy       5940 CATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACA 5999
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       5985 CATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTTCTTTTCTCGAAGCATGACA 6044

Qy       6000 AGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGTGCTGGGTTAT 6058
              ||||||| |||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       6045 AGTGTAT-GCAATACTTACATGCTCACTCGTTTACCTGTGCTTAGTTTGTGCTGGGTTAT 6103

Qy       6059 TTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTAC 6118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6104 TTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTAC 6163

Qy       6119 ACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAG 6174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6164 ACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAG 6219
```

```
RESULT 4
AB052621
LOCUS       AB052621                6222 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36E mRNA, complete cds.
ACCESSION   AB052621
VERSION     AB052621.1  GI:11862940
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6222)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
```

```
                     /clone="DDM36E"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)-vector:pDDM36EORF"
     gene            1..6222
                     /gene="DDM36E"
     CDS             50..3811
                     /gene="DDM36E"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36E"
                     /protein_id="BAB19279.1"
                     /db_xref="GI:11862941"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVH
                     NQSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGG
                     RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPT
                     PDLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLV
                     TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                     SPLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLE
                     SDTRYFFKMGARTEVGPGPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLA
                     CMCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWS
                     PPPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRP
                     ALTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGC
                     PDLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEA
                     ASASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1321 a   1772 c   1653 g   1476 t
ORIGIN

  Query Match             97.6%;  Score 6025.8;  DB 10;  Length 6222;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 6136;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    8;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 109

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 169

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 229

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 289

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 349

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 409

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 469

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 529

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 589

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
              ||||||||||||||| ||||||||||||||||||||||  |  |||||||||||||||||
Db        590 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 649

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 709

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 769

Qy        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        770 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTA 829

Qy        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        830 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 889

Qy        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 949

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| 
Db        950 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGAT 1009

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1010 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCC 1069

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1070 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1129

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1130 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1189

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1190 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGC 1249

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1250 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1309

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1310 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1369

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1370 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1429

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1430 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1489

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1490 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1549

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1550 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1609
```

```
Qy   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1669

Qy   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1729

Qy   1681 TACAAGATAGAGTACGGTTTGGGGAAG---GAAGATCAGGTTTTCTCCACCGAGGTGCCT 1737
          |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db   1730 TACAAGATAGAGTACGGTTTGGGGAAGGAAGAAGATCAGGTTTTCTCCACCGAGGTGCCT 1789

Qy   1738 GGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGG 1797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 GGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGG 1849

Qy   1798 ATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACA 1857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1850 ATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACA 1909

Qy   1858 CCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCA 1917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 CCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCA 1969

Qy   1918 AAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGA 1977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 AAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGA 2029

Qy   1978 TACAAACTCTACTGGGGAGAGGTGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCA 2037
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   2030 TACAAACTCTACTGGAGAGAGGTGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCA 2089

Qy   2038 GGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAG 2097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2090 GGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAG 2149

Qy   2098 CAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTC 2157
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   2150 CAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTC 2209

Qy   2158 AACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACG 2217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2210 AACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACG 2269

Qy   2218 CCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCA 2277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2270 CCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCA 2329

Qy   2278 AACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATT 2337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2330 AACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATT 2389

Qy   2338 GTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTAC 2397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2390 GTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTAC 2449

Qy   2398 TATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAG 2457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2450 TATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAG 2509

Qy   2458 TTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGC 2517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2510 TTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGC 2569

Qy   2518 TCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACA 2577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2570 TCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACA 2629

Qy   2578 CCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAG 2637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2630 CCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAG 2689

Qy   2638 TATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACC 2697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2690 TATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACC 2749

Qy   2698 ACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTC 2757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2750 ACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTC 2809

Qy   2758 TTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTG 2817
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2810 TTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTG 2869

Qy   2818 ATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATC 2877
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   2870 ATTACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATC 2929

Qy   2878 GTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAA 2937
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2930 GTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACGA 2989

Qy   2938 AGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCG 2997
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   2990 AGCTCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCG 3049

Qy   2998 CTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCC 3057
          ||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||||
Db   3050 CTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCC 3109

Qy   3058 CTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCT 3117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3110 CTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCT 3169

Qy   3118 GAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAG 3177
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3170 GAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAG 3229

Qy   3178 ATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTG 3237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3230 ATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTG 3289

Qy   3238 CCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACC 3297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3290 CCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACC 3349

Qy   3298 GGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAG 3357
          ||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||||||
Db   3350 GGGCAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAG 3409

Qy   3358 AAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTC 3417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3410 AAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTC 3469

Qy   3418 GGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTG 3477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 GGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTG 3529

Qy   3478 ACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGG 3537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3530 ACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGG 3589

Qy   3538 CTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGCCAGAGAGGCTCACCTGC 3597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3590 CTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGCCAGAGAGGCTCACCTGC 3649

Qy   3598 TTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATA 3657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3650 TTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATA 3709

Qy   3658 GAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCA 3717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3710 GAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCA 3769

Qy   3718 AGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGC 3777
```

```
Db   3770 AGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGC 3829

Qy   3778 TCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGG 3837
Db   3830 TCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGG 3889

Qy   3838 AGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCA 3897
Db   3890 AGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCA 3949

Qy   3898 TCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTG 3957
Db   3950 TCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTG 4009

Qy   3958 TGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACT 4017
Db   4010 TGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACT 4069

Qy   4018 GTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGA 4077
Db   4070 GTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGA 4129

Qy   4078 GCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTC 4137
Db   4130 GCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTC 4189

Qy   4138 TACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTC 4197
Db   4190 TACAAAGACAGCGCCAGCCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTC 4249

Qy   4198 CACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAG 4257
Db   4250 CACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAG 4309

Qy   4258 GTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTA 4317
Db   4310 GTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTA 4369

Qy   4318 ATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCT 4377
Db   4370 ATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCT 4429

Qy   4378 GCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGG 4437
Db   4430 GCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGG 4489

Qy   4438 GCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCC 4497
Db   4490 GCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCC 4549

Qy   4498 ATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTA 4557
Db   4550 ATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTA 4609

Qy   4558 ACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCT 4617
Db   4610 ACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCT 4669

Qy   4618 CCCCTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCC 4677
Db   4670 CCCCTGCCAACCGCCTTCCTGGGCAGATGA-CCCACTCCAAGTAGATCTGGAAAAGA-CC 4727

Qy   4678 TTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGA 4737
Db   4728 TTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGA 4787

Qy   4738 GATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGC 4797
Db   4788 GATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGC 4847

Qy   4798 TGAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCA 4857
Db   4848 TG-AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCA 4906

Qy   4858 GTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGT 4917
Db   4907 GTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGT 4966

Qy   4918 TCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGC 4977
Db   4967 TCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGC 5026

Qy   4978 TGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTT 5037
Db   5027 TGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTT 5086

Qy   5038 ACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGT 5097
Db   5087 ACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGT 5146

Qy   5098 CACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGAT 5157
Db   5147 CACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGAT 5206

Qy   5158 CCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGT 5217
Db   5207 CCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGT 5266

Qy   5218 GTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATT 5277
Db   5267 GTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATT 5326

Qy   5278 TTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTT 5337
Db   5327 TTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTT 5386

Qy   5338 ATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTC 5397
Db   5387 ATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTC 5446

Qy   5398 AGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAG 5457
Db   5447 AGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAG 5506

Qy   5458 GATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATG 5517
Db   5507 GATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATG 5566

Qy   5518 TTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCC 5577
Db   5567 TTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTTTTTTCC 5626

Qy   5578 TCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGT 5637
Db   5627 TCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGT 5686

Qy   5638 CATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGC 5697
Db   5687 CATTCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGC 5744

Qy   5698 AAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAA 5756
Db   5745 AAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTTCTCAAAAGCAAATAA 5804

Qy   5757 GTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACT 5816
Db   5805 GTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACT 5864

Qy   5817 AGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAG 5876
Db   5865 AGCATTATATATGATTTGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAG 5924

Qy   5877 CACCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTT 5936
Db   5925 CAGCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTT 5984
```

```
Qy       5937 TGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATG 5996
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       5985 TGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTTCTTTTCTCGAAGCATG 6044

Qy       5997 ACAAGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGTGCTGGGT 6055
              |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db       6045 ACAAGTGTAT-GCAATACTTACATGCTCACTCGTTTACCTGTGCTTAGTTTGTGCTGGGT 6103

Qy       6056 TATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGT 6115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6104 TATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGT 6163

Qy       6116 TACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTCCAG 6174
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6164 TACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTCCAG 6222


RESULT 5
AK122535
LOCUS       AK122535              6301 bp    mRNA    linear   ROD 15-MAR-2003
DEFINITION  Mus musculus mRNA for mKIAA1628 protein.
ACCESSION   AK122535
VERSION     AK122535.1  GI:28972801
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
            Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: II. The complete nucleotide sequences of 400 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10, 35-48 (2003)
REFERENCE   2  (bases 1 to 6301)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1..6301
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbg00013"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1..6301
                     /gene="mKIAA1628"
     CDS             <1..3887
                     /gene="mKIAA1628"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=3
                     /evidence=not_experimental
                     /product="mKIAA1628 protein"
                     /protein_id="BAC65817.1"
                     /db_xref="GI:28972802"
                     /translation="EGSQENPSAWHGRGELPLPQETTVKLSCDEGPLQVILGPEQAVV
                     LDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRI
                     WKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGL
                     PAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSARQRFSQEAS
                     LTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSWVRQGDQSK
                     KELGIQTSGCDAALQGCSGNLVVGEGVEGTVPASPSSLLLLCHQMESLSPRMSSFWAG
                     PIYSSPARSLGTLESMSAEPTSPARVISPLRLLSSECLLPQPSRRRPRRSRGRGPAPR
                     ASCAGRPGSHGPRCTGCTTGSRCDPMGRVKVQGGGGSLVITQIGLQDAGYYQCVAENS
                     AGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQK
                     ARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDV
                     PSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEEDQVFSTEVPGNE
                     TQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAK
                     MESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKV
                     KQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVH
                     AESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKP
                     FTKYEFAVQSHGVDMDGPPGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTE
                     PNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYPFKMGARTEVGP
                     GPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLACMCAGLRRSSHREALPG
                     LSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSPPPSDVEDKAEVHSLMG
                     GSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPALTRALLPPAGTGQTLL
                     LQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCPDLHLQDLEPEEPLTAE
                     TLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAASASCSCSDLQPSTAIE
                     EAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1362 a   1770 c   1669 g   1500 t
ORIGIN

  Query Match             94.1%;  Score 5810.2;  DB 10;  Length 6301;
  Best Local Similarity   97.0%;  Pred. No. 0;
  Matches 6077;  Conservative    0;  Mismatches   33;  Indels  157;  Gaps    9;

Qy         62 GCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCC 121
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         40 GAGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCC 99

Qy        122 TGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAG 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 TGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAG 159

Qy        182 CTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATG 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 CTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATG 219

Qy        242 AGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 AGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAG 279

Qy        302 ACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTC 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 ACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTC 339

Qy        362 TGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACAC 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACAC 399

Qy        422 TCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCT 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCT 459

Qy        482 TTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGA 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGA 519

Qy        542 CCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATG 601
              |||||||||||||| ||||||||||||||||||||||  |  ||||||||||||||||||
Db        520 CCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATG 579

Qy        602 TCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGAT 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGAT 639

Qy        662 TCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 TCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGC 699

Qy        722 AGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAG 781
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        700 AGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAG 759
```

```
Qy    782 TGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAG---- 837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 TGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGGGG 819

Qy    838 ------------------------------------------------------------ 837

Db    820 ATCAGAGTAAGAAGGAACTGGGGATTCAAACCAGCGGCTGTGATGCAGCTTTGCAGGGAT 879

Qy    838 ------------------------------------------------------------ 837

Db    880 GTAGTGGGAACCTCGTAGTTGGAGAGGGTGTAGAGGGGACAGTTCCAGCGTCCCCCTCCA 939

Qy    838 ----------------------GATGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGG 875
                                ||||||||||||||||||||||||||||||||||||||
Db    940 GCCTGCTCCTTCTCTGTCACCAGATGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGG 999

Qy    876 CCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCG 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1000 CCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCG 1059

Qy    936 AGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGC 995
          |||||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||
Db   1060 AGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGC 1119

Qy    996 TGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCG 1055
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1120 TGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCG 1179

Qy   1056 CTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGAT 1115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1180 CTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGAT 1239

Qy   1116 CCCGTTGCGACCCAATGGG-CGCGTCAAGGTGCAGGCGGTGGCGGCAGCTTGGTCATCA 1174
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1240 CCCGTTGCGACCCAATGGGCCGCGTCAAGGTGCAGGCGGTGGCGGCAGCTTGGTCATCA 1299

Qy   1175 CTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAA 1234
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1300 CTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAA 1359

Qy   1235 CTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGA 1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1360 CTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGA 1419

Qy   1295 CTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTG 1354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1420 CTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTG 1479

Qy   1355 AGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGG 1414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1480 AGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGG 1539

Qy   1415 ACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACC 1474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1540 ACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACC 1599

Qy   1475 TGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCC 1534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1600 TGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCC 1659

Qy   1535 GAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGC 1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1660 GAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGC 1719

Qy   1595 TTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCA 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1720 TTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCA 1779

Qy   1655 GCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAG---GAAG 1711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db   1780 GCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGAAG 1839

Qy   1712 ATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTC 1771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1840 ATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTC 1899

Qy   1772 AGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCC 1831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1900 AGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCC 1959

Qy   1832 CTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTG 1891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1960 CTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTG 2019

Qy   1892 CCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGC 1951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2020 CCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGC 2079

Qy   1952 CCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGG 2011
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   2080 CCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGAGG 2139

Qy   2012 AGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGC 2071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2140 AGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGC 2199

Qy   2072 CCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGC 2131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 CCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGC 2259

Qy   2132 CGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGA 2191
           |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   2260 TGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGA 2319

Qy   2192 AGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGC 2251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2320 AGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGC 2379

Qy   2252 CTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGA 2311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2380 CTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGA 2439

Qy   2312 AGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGC 2371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2440 AGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGC 2499

Qy   2372 TCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCG 2431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2500 TCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCG 2559

Qy   2432 GCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATG 2491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2560 GCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATG 2619

Qy   2492 GGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTT 2551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2620 GGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTT 2679

Qy   2552 CTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCA 2611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2680 CTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCA 2739

Qy   2612 CGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGC 2671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2740 CGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGC 2799

Qy   2672 CCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATG 2731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2800 CCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATG 2859

Qy   2732 GCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTG 2791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2860 GCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTG 2919

Qy   2792 GGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGG 2851
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
```

```
Db   2920 GGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAAAGACATTCTCAGACTCCTTGG 2979

Qy   2852 ATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGG 2911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2980 ATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGG 3039

Qy   2912 CCTGCATGTGTGCTGGCCTACGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCT 2971
          |||||||||||||||||||||||| ||||||||||||||||||||| |||||||||||||
Db   3040 CCTGCATGTGTGCTGGCCTACGACGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGTCCT 3099

Qy   2972 CCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTG 3031
          ||||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db   3100 CCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGTG 3159

Qy   3032 TCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCAC 3091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3160 TCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCAC 3219

Qy   3092 CACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAG 3151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3220 CACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAG 3279

Qy   3152 ATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGG 3211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3280 ATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGG 3339

Qy   3212 CAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCC 3271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3340 CAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCC 3399

Qy   3272 GTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATG 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   3400 GTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTATATG 3459

Qy   3332 ACGGCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAG 3391
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3460 ACGCCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAG 3519

Qy   3392 CTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCC 3451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3520 CTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCC 3579

Qy   3452 AAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTG 3511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3580 AAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTG 3639

Qy   3512 TGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGAGGGTGCCAACCAACAA 3571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3640 TGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGAGGGTGCCAACCAACAA 3699

Qy   3572 CCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCT 3631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3700 CCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCT 3759

Qy   3632 CAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAG 3691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3760 CAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAG 3819

Qy   3692 CCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGG 3751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3820 CCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGG 3879

Qy   3752 TCCCCTGAGCAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTG 3811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3880 TCCCCTGAGCAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTG 3939

Qy   3812 TACTAGAGATATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAG 3871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3940 TACTAGAGATATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAG 3999

Qy   3872 AACTTTACTCTCCCTCATATTCTGCATCACATACAGGAGGACTTGAGACACAGCTCTGT 3931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4000 AACTTTACTCTCCCTCATATTCTGCATCACATACAGGAGGACTTGAGACACAGCTCTGT 4059

Qy   3932 GTAATGGACACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGG 3991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4060 GTAATGGACACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGG 4119

Qy   3992 AAACCTCTCCCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGT 4051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4120 AAACCTCTCCCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGT 4179

Qy   4052 AACAGTGATCAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACAC 4111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4180 AACAGTGATCAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACAC 4239

Qy   4112 TCTGAGAGTGTCTCCAGTGCTGTGTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCT 4171
          |||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db   4240 TCTGAGAGTGTCTCCAGTGCTGTGTCTACAAAGACAGCGCCAGCCCCTCTTCTGTCAGCT 4299

Qy   4172 GTGCTGAGCAGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCAT 4231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4300 GTGCTGAGCAGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCAT 4359

Qy   4232 GGCTGTCATCCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCC 4291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4360 GGCTGTCATCCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCC 4419

Qy   4292 CTCTCCTTTGAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAA 4351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4420 CTCTCCTTTGAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAA 4479

Qy   4352 TCTTCTTTCGCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTC 4411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4480 TCTTCTTTCGCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTC 4539

Qy   4412 ATGTCATCTTGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCT 4471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4540 ATGTCATCTTGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCT 4599

Qy   4472 CACTGCTGTTTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCT 4531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4600 CACTGCTGTTTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCT 4659

Qy   4532 GAGTTCAACCTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCAT 4591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4660 GAGTTCAACCTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCAT 4719

Qy   4592 AGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGAGCCC 4651
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   4720 AGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGA-CCC 4778

Qy   4652 ACTCCAAGTAGATCTGGAAAAGACCCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTT 4711
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   4779 ACTCCAAGTAGATCTGGAAAAGA-CCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTT 4837

Qy   4712 GAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGA 4771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4838 GAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGA 4897

Qy   4772 AGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAG 4831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4898 AGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAG 4957

Qy   4832 AGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGG 4891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4958 AGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGG 5017

Qy   4892 AGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGA 4951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5018 AGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGA 5077

Qy   4952 GTAGCAAAGAGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGC 5011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5078 GTAGCAAAGAGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGC 5137
```

```
Qy       5012 TGGCTCATTTTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGG 5071
Db       5138 TGGCTCATTTTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGG 5197

Qy       5072 CCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACC 5131
Db       5198 CCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACC 5257

Qy       5132 TCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTC 5191
Db       5258 TCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTC 5317

Qy       5192 TATTTCAACTGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAA 5251
Db       5318 TATTTCAACTGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAA 5377

Qy       5252 CCAGTCTATTTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGT 5311
Db       5378 CCAGTCTATTTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGT 5437

Qy       5312 TGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCAC 5371
Db       5438 TGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCAC 5497

Qy       5372 CCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATG 5431
Db       5498 CCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATG 5557

Qy       5432 CTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAA 5491
Db       5558 CTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAA 5617

Qy       5492 GTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATT 5551
Db       5618 GTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATT 5677

Qy       5552 TTGTTTTCCTGGTTAGATTCTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGT 5611
Db       5678 TTGTTTTCCTGGTTAGATTTTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGT 5737

Qy       5612 AGGCTCTTAGCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAG 5671
Db       5738 AGGCTCTTAGCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACT--GA 5795

Qy       5672 GTTCTGGGTTTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTG 5731
Db       5796 GGTTCTGGTTTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTG 5855

Qy       5732 GAAAATTG-TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGT 5790
Db       5856 GAAAATTGTTTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGT 5915

Qy       5791 TGATAAGAAAGTAGCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGG 5850
Db       5916 TGATAAGAAAGTAGCCAGAACTCACTAGCATTATATATGATTTGGGTTCTGAGTAACTGG 5975

Qy       5851 GGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTCTTATTAGGAAAGTCTGTTGGCCTTT 5910
Db       5976 GGAGTGTTAGCTTTGTGACTTTGTAGCAGCAGGTCTTATTAGGAAAGTCTGTTGGCCTTT 6035

Qy       5911 TACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCAT 5970
Db       6036 TACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCAT 6095

Qy       5971 TTAAGAATTCCTTTTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGT 6030
Db       6096 TTAAGAATTTCTTTTCTCGAAGCATGACAAGTGTAT-GCAATACTTACATGCTCACTCGT 6154

Qy       6031 TTACCTG-GCTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTC 6089
Db       6155 TTACCTGTGCTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTC 6214

Qy       6090 CTTACAAATATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGT 6149
Db       6215 CTTACAAATATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGT 6274

Qy       6150 AAAGAATTAATAAACTATTTTCCAGAC 6176
Db       6275 AAAGAATTAATAAACTATTTTCCAGAC 6301
```

```
RESULT 6
AX191303
LOCUS       AX191303                3756 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 45 from Patent WO0149714.
ACCESSION   AX191303
VERSION     AX191303.1  GI:15209554
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 45 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..3756
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
BASE COUNT      774 a    1167 c    1105 g     710 t
ORIGIN

  Query Match             60.8%;  Score 3756;  DB 6;  Length 3756;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3756;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
```

```
Qy       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
              ||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 7
AB052622
LOCUS       AB052622                6485 bp    mRNA    linear   PRI 16-MAR-2002
DEFINITION  Homo sapiens hDDM36 mRNA, complete cds.
ACCESSION   AB052622
VERSION     AB052622.1  GI:19570397
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6485)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6485
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="hDDM36"
                     /cell_line="NHOst"
                     /cell_type="osteoblast"
                     /note="similar to mouse DDM36 and Nope
                     (NM_020043)-vector:phDDM36"
     gene            1..6485
                     /gene="hDDM36"
     CDS             211..3963
                     /gene="hDDM36"
                     /note="similar to mouse DDM36 and Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="hDDM36"
                     /protein_id="BAB86306.1"
                     /db_xref="GI:19570398"
                     /translation="MARGDAGRGRGLLALTFCLLAARGELLLPQETTVELSCGVGPLQ
                     VILGPEQAAVLNCSLGAAAAGPPTRVTWSKDGDTLLEHDHLHLLPNGSLWLSQPLAPN
                     GSDESVPEAVGVIEGNYSCLAHGPLGVLASQTAVVKLATLADFSLHPESQTVEENGTA
                     RFECHIEGLPAPIITWEKDQVTLPEEPRLIVLPNGVLQILDVQESDAGPYRCVATNSA
                     RQHFSQEALLSVAHRGSLASTRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVS
                     WVRQDGKPISTDVIVLGRTNLLIANAQPWHSGVYVCRANKPRTRDFATAAAELRVLAA
                     PAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNGAPLRPNGRVKVQGGGGSLVI
                     TQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPSAPTRVTATPLSSSAVLVAWE
                     RPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQ
                     LGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVAWLPLPPSLSNGQVVKYKIEY
                     GLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAGTAAGFGAPSQWMHHRTPSMH
                     NQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYKLYWREVGAEEEANGDRLPGG
                     RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPA
                     PDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFSPWGLRNASLV
                     TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                     SPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEHQWTLLTTQGNIFSAEVHGLE
                     SDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLDMHSVTGIIVGVCLGLLCLLA
                     CMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPPSPPAAHELESLVHPHPQDWS
                     PPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSGLSWAGSWAGCELPQAGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACRNQVEAEVIVHSDFSASNGNP
                     DLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDRELGGCELAAPGPDRLTCLPEAA
                     SASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLPRSPVSSSA"
BASE COUNT     1293 a   1926 c   1772 g   1494 t
ORIGIN

  Query Match             55.0%;  Score 3396.6;  DB 9;  Length 6485;
  Best Local Similarity   75.6%;  Pred. No. 0;
  Matches 4784;  Conservative    0;  Mismatches 1324;  Indels  223;  Gaps   37;
```

```
Qy      3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62
Db    219 GGGGGACGCCGGCCGCGGCCGTGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 278

Qy     63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122
Db    279 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACT 338

Qy    123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182
Db    339 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 398

Qy    183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242
Db    399 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 458

Qy    243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302
Db    459 CCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 518

Qy    303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362
Db    519 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGTCATTGAAGGCAACTATTCGTGCCT 575

Qy    363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422
Db    576 AGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACT 635

Qy    423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482
Db    636 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 695

Qy    483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542
Db    696 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 755

Qy    543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602
Db    756 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 815

Qy    603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662
Db    816 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 875

Qy    663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722
Db    876 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCA 935

Qy    723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782
Db    936 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 995

Qy    783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842
Db    996 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGG 1055

Qy    843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902
Db   1056 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 1115

Qy    903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962
Db   1116 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 1175

Qy    963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGA 1022
Db   1176 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGA 1235

Qy   1023 GGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCC 1082
Db   1236 GGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCC 1295

Qy   1083 ACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAA 1142
Db   1296 GCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAA 1355

Qy   1143 GGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTA 1202
Db   1356 GGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTA 1415

Qy   1203 CTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGT 1262
Db   1416 CTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGT 1475

Qy   1263 AGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAG 1322
Db   1476 GGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAG 1535

Qy   1323 CTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTT 1382
Db   1536 CTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTT 1595

Qy   1383 CTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAA 1442
Db   1596 CTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAA 1655

Qy   1443 CAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTA 1502
Db   1656 CAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTA 1715

Qy   1503 CGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATAC 1562
Db   1716 CGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACAC 1775

Qy   1563 ACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGA 1622
Db   1776 ACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGA 1835

Qy   1623 CATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTA 1682
Db   1836 CATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTA 1895

Qy   1683 CAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAA 1742
Db   1896 CAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAA 1955

Qy   1743 TGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTC 1802
Db   1956 TGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTC 2015

Qy   1803 AGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGG 1862
Db   2016 GGCTGGTACAGCAGCCGGCTTCGGGGCCCCTCCCAGTGGATGCATCACAGGACGCCCAG 2075

Qy   1863 TGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGAT 1922
Db   2076 TATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGAT 2135

Qy   1923 GGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAA 1982
Db   2136 GGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAA 2195

Qy   1983 ACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGG 2042
Db   2196 ACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGG 2255

Qy   2043 TCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTA 2102
Db   2256 CCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTA 2315

Qy   2103 TGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAA 2162
Db   2316 TGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAA 2375

Qy   2163 ACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGA 2222
```

```
Db   2376 ACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGA 2435

Qy   2223 CCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAG 2282
Db   2436 CATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAG 2495

Qy   2283 CTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAA 2342
Db   2496 CTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAA 2555

Qy   2343 CTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATAC 2402
Db   2556 CTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACAC 2615

Qy   2403 CAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGC 2462
Db   2616 CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGC 2675

Qy   2463 GGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCAC 2522
Db   2676 AGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCAC 2735

Qy   2523 CCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATC 2582
Db   2736 CCTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTC 2795

Qy   2583 CACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCT 2642
Db   2796 CACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCT 2855

Qy   2643 AATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGA 2702
Db   2856 GATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCA 2915

Qy   2703 GGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAA 2762
Db   2916 GGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAA 2975

Qy   2763 GATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTAC 2822
Db   2976 GATGGGGGCGCGCACAGAGGTGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCAC 3035

Qy   2823 TCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTGGG 2882
Db   3036 GCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATCATCGTGGG 3095

Qy   2883 TGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGCTC 2942
Db   3096 TGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCC 3155

Qy   2943 CCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTCTA 3002
Db   3156 CCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTA 3215

Qy   3003 CACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGT 3062
Db   3216 CTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGT 3275

Qy   3063 GCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGT 3122
Db   3276 GCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGT 3335

Qy   3123 ACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTC 3182
Db   3336 GCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGATCTC 3395

Qy   3183 CTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCA 3242
Db   3396 CTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCA 3455

Qy   3243 GGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCA 3302
Db   3456 GG---CAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCA 3512

Qy   3303 GACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAGCC 3362
Db   3513 GACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGAGGAAGAAGTC 3572

Qy   3363 GTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGC 3422
Db   3573 ACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGC 3632

Qy   3423 ATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGC 3482
Db   3633 ATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCC 3692

Qy   3483 AGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGG 3542
Db   3693 AGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGA 3752

Qy   3543 CAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCC 3602
Db   3753 CAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCC 3812

Qy   3603 AGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGA 3662
Db   3813 AGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGA 3872

Qy   3663 GGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGCCT 3722
Db   3873 GACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCT 3932

Qy   3723 TCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGA---------------AG 3766
Db   3933 GCCCAGATCCCCGGTCTCCTCCTCTGCCTAGCTCTTCCCAGAGGATGTGGTTTGGGGCAG 3992

Qy   3767 GCAGATATGGCTCAGGAA------------CATGCCATGCATGGCTACACATGTGTGTA 3813
Db   3993 GCAGGTATGGATCACATAGGATGCGATACCTGTGGCCGTGTATGTCCACATGTGTGCCTG 4052

Qy   3814 CTAGAGATATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAA 3873
Db   4053 TAGATACATCATCAAGCCCTTTGGAGCTTCCTAAGTTGCTTTGGCTGAGGGGAGAGGAAA 4112

Qy   3874 ------CTTTACTCTCCCTCATATTCTGCATCACATACAGGAGGGACTTGAGACACAGCT 3927
Db   4113 ACATGGATTATTCACTCCCCCCATACTCTTTGTGATACACATGTGACATGTGAAAGACAT 4172

Qy   3928 CTGTGTAATGGACACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTG-----TGCTGGT 3982
Db   4173 ACGAGACATAGCTACATGTGATGTGCACATGTGTGAAGTGCATGTATGCGTACTGGTTGT 4232

Qy   3983 TGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCA 4042
Db   4233 TGAGCTGGGAAACCGTGGCCCAGGCAGTGGTCACTACAGCCTGATTGGTCCTCCAGGTCA 4292

Qy   4043 GGATGGTGTA--ACAGTGATCAGTG-CCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGC 4099
Db   4293 GAACGGTGCCCCACAGTGGTCAGTCCCCAGCCCTGTGGGCCCCCACCTCCATCGCCCAGC 4352

Qy   4100 CTTTTATT--ACACTCTGAGAGTGTCTCCAGTGCTGTGTCTACAAAGACAGCGCCCAGCC 4157
Db   4353 CTTTTATTACACACTCTGAGAGTGTCTCCAATGCCTGTCTGACAAAGACAGTCCCAGCCC 4412

Qy   4158 CTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCA 4217
Db   4413 ATTCTC---CTGTCTGGCTGGGTTGGGTGCAA-----------------------GCA 4444

Qy   4218 GCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAAC 4277
Db   4445 GGCTCTGAATGCCTGGCATTTCAGCTGCATCACCTCCCAGCTCCTTATTGCCCAAATAGA 4504

Qy   4278 CCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTG 4337
Db   4505 GAGGGTGGCCCTGGCTCCCCTCCGAGCAACTCTGCATTTAATTTTGTAATCTGGGAAGTG 4564
```

```
Qy       4338 CCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTC 4397
Db       4565 CCTGGTTTTGAAAATCCGCTTTCTCTCACTCTTCCCCTCCTTCCTTGCCCCTGGCTGCTC 4624

Qy       4398 TGGTGATCTGTCT-CATGTCATCTTGCTCGATTATC-CTGGGGCCCTTCTCTT--TCCCA 4453
Db       4625 TAGTGTTCTGTCTCCCAGTCACCTCGCTCTCCCAGCACCAGTGCCCTTCTCCTGCTCCCA 4684

Qy       4454 TGATGCCCCTGATTTCCTCACTGCTGTTTTCATT---------TCTGTCTGCCATGCTTG 4504
Db       4685 GATACTCTTTCCTTTCCTCTCTCCTGTTTTCCTTCCTCTGCTATCTCTCACACCTCTCCC 4744

Qy       4505 TCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACAACAT 4564
Db       4745 AGACTATGTCATCTTGTTCTCCTGCCTGGGTTCAAACTCTGCATCCTTCTCTAACAACGT 4804

Qy       4565 GACTACCTCATGTCTGCTTCA---------GACCATAGTGTGACCCCTGGGTCCCCACAG 4615
Db       4805 GACTACCTCATGTCTGCTTCAAGGCCCCCGTGCCCTTCCTGTATCCGCGGCTGCCGCGCA 4864

Qy       4616 CTCCCCTGCCAACCGCCTTCCT---------------------------------GGGC 4641
Db       4865 CTCGCCTGCCATCCTCCTGCCTCCTCTTCACTCAGTGCTTCTGCTTGCCCTGCCCCAGGC 4924

Qy       4642 AGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTGTGGCTTGTCTGGCTGCCCTCC 4701
Db       4925 AGCCCACCCACGCCCAGTGCGGGTGTGGAGAAGATCTTCTGGCTTCCCTGCAT--CTTGC 4982

Qy       4702 CCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGA 4761
Db       4983 CTTTGGGATTGGGATCCAAGGGTTCTCCATGGATG----GATCCAAGTCATAGAGGGGAA 5038

Qy       4762 ACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCC-- 4819
Db       5039 TGTTTGAGACAGGGAAGGGGGCTGTGATCCAGAGGCTCAGAATAAAAAGATGCCCTCCCT 5098

Qy       4820 -ATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGG 4878
Db       5099 TCTATGCAGGGGGGCAAGTTTACTGGATGGAGATGATTTGGGCCTCTCTTCCAGAAGAAG 5158

Qy       4879 GTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGA 4938
Db       5159 CTAAAGGAAGAGAAGGGGAGTGAGAGTTCAGGGAGGCCCTTCCCACCCTGTGAGGCTTGA 5218

Qy       4939 ACCTGGCTTCAGAGTAGCAAAGAGTTT------CCTCCAAGATGCTGTAAGGGAAGTCTT 4992
Db       5219 CTTGATCTGGATTGGGGATGACAGGAATCTCACCCTCTGGGGTGCTGGCAAGGAGGTCTT 5278

Qy       4993 TGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACACTGAATCCCAAA 5052
Db       5279 TGCACAGGAAAAGG--GGTAGCTCATTTCAGTTTGTTTTTTCTTTAAATTGAATCCTCAA 5336

Qy       5053 ATCAT----------CTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCACAA 5102
Db       5337 GTCATTTTCTGTTCACCTGCCGCACAGGGACAAGCTTGACTTCTATTTTCTGTGTAGTGA 5396

Qy       5103 GAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCAGT 5162
Db       5397 AAACAATGTCATTTATTTGGTTTTTCACCTCAGCCCTCTCATAGGAGCATAGAATGTTAG 5456

Qy       5163 GGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTTAA 5222
Db       5457 GGTCTTTACTCCCTA-----------------ATGATGTCTGATTGGCACATCAAGAGTTAA 5501

Qy       5223 CTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTCCC--TTTATTTTATTTTA 5280
Db       5502 CTCTGCCTTCTGGGCCAAATTCGAAATAACCAGTCCATTTTTCCTTTTTTTTTTTTTTTT 5561

Qy       5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTT---------CTGAAAGCGTCT 5331
Db       5562 TTTTAAATGGTGGAATGTCTCTCAGCACAGTTGCGGCTTCCTCAAACCCTGAAAGCATCT 5621

Qy       5332 GTGTTTATTAGCCTTGTGTGTCACTCATGTTTGA---CCCCACCCACATTTCCTTCTCCT 5388
Db       5622 GTGTTTATTATACTCGGGTGTCACTCACTGTTGATGTCTGCACCTACGTTTCCACCTCCT 5681

Qy       5389 CCCC--TCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTC 5446
Db       5682 CCCCCTCCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGTTGGTTTTGTATCTC 5741

Qy       5447 ATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAAC 5506
Db       5742 GTTAAAGACAGAATTGTCACTTGTAGTATTTCTGTAGCATTCAGCGCTGCTGTGGCTAAC 5801

Qy       5507 ACCAC--CGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGT 5564
Db       5802 ACCACTGTGTATGTTTCATCATTGCTCTGAAGGTCAAAAGCCTCATTTTATTTTGCTGGT 5861

Qy       5565 TAGATTCTTTTCCT----------------CCTTGCCTTGAATGAAATAACCGTTTTAAC 5608
Db       5862 TTGATTTTTTTTTTTAAAGAAGAAAAAAAAACTGCCCTGAATTAAATGGCTGTTTTAAC 5921

Qy       5609 AGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTT 5668
Db       5922 AGTAGGCTCTTAGCATTATACCACATAGTCATTTTCATGTTCTTGTTTAACAGGCACT- 5980

Qy       5669 GAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGG 5728
Db       5981 -GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGTTGG 6039

Qy       5729 GTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCAC 5787
Db       6040 GTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCAC 6099

Qy       5788 AGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAAC 5847
Db       6100 AGTTGATAAGAAAGTGGCCATTTCTCACTAGCACTATATATGATTTGGGCTCTGGGTAAT 6159

Qy       5848 TGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTCT-TATTAGGAAAGTCTGTTGGC 5906
Db       6160 TTGGAAGTGTTAGGTTTGTGTCTTTGTAGCAGTATTTTTATTAGAAAAGAATCTATTGGC 6219

Qy       5907 CTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTG-GAGT 5965
Db       6220 CTTTTACAGGGTATTAATCCCTTTGTCACCTACCATTGATGCCTTAAGTTTTCTGAGTCT 6279

Qy       5966 CTCATTTAAGAATTCCTTTTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCA 6025
Db       6280 CAATTAAAAATCTTCCTTTTCTTGATGCATGACAAGTGT-----AATCAGTACTTGCTCA 6334

Qy       6026 CTCGTTTACCTG-GCTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCT 6084
Db       6335 TTTATTTGTCTGTATTTAGTTTATGCTGTACTATTTAATTATCCTTCCAGCGTTTTTTTT 6394

Qy       6085 TCCTCCTTACAAATATGATATTCTTTATTGTTACACTAAGGTGTTGA-TCATGTATCTGT 6143
Db       6395 TTCTCCTTACAAATATGATACTCTTTAGTGTTAAGCTAAGGCATTGATTCATGTATCTGT 6454

Qy       6144 CCCTGTAAAGAATTAATAAACTATTTTCCAG 6174
Db       6455 CCTTATAATGAATTAATAAACTATTTTCCAG 6485
```

```
RESULT 8
AX191261
LOCUS       AX191261                2796 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 3 from Patent WO0149714.
ACCESSION   AX191261
VERSION     AX191261.1  GI:15209512
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
```

```
TITLE      Nope polypeptides, encoding nucleic acids and methods of use
JOURNAL    Patent: WO 0149714-A 3 12-JUL-2001;
           Neurosciences Research Foundation Inc. (US)
FEATURES            Location/Qualifiers
     source         1. .2796
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
     CDS            <1. .>2796
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="CAC51121.1"
                    /db_xref="GI:15209513"
                    /translation="GELPLPQETTVKLSCDEGPLQVILGPEQAVVLDCTLGATAAGPP
                    TRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRIWKVTEGSYSCLAH
                    SPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGLPAPIITWEKDQVT
                    VPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSARQRFSQEASLTVALRGSLEATR
                    GQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSWVRQDGKPISTDVIVLGRTNLL
                    IASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAPAISQAPEALSRTRASTARFVC
                    RASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGTA
                    CAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQKARGV
                    DNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDVPSAA
                    PQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEDQVFSTEVPGNETQLTL
                    NSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAKMESLV
                    VSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKVKQYEL
                    TQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVHAESNS
                    STSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKPFTKYE
                    FAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEI
                    VEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSR
                    LQDVITLQETFSDSLDVH"
BASE COUNT      591 a    849 c    821 g    535 t
ORIGIN

  Query Match             45.3%;  Score 2796;  DB 6;  Length 2796;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy       2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy       2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy       2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy       2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy       2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy       2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy       2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy       2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy       2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy       2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy       2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859
              ||||||||||||||||||||||||||||||||||||
Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

```
RESULT 9
AC112161/c
LOCUS       AC112161               115135 bp    DNA     linear   ROD 21-SEP-2002
DEFINITION  Mus Musculus chromosome 9 BAC clone MGS1-117K9 ES cell line,
            Complete Sequence, complete sequence.
ACCESSION   AC112161
VERSION     AC112161.9  GI:21426127
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 115135)
  AUTHORS   Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 115135)
  AUTHORS   Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   3  (bases 1 to 115135)
  AUTHORS   Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   4  (bases 1 to 115135)
  AUTHORS   Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   5  (bases 1 to 115135)
  AUTHORS   Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On Jun 14, 2002 this sequence version replaced gi:21358700.
            ------------------Genome Center:
            Center:    Harvard Partners Genome Center
            Center Code:    HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact:   hpgc@mendel.mgh.harvard.edu

            CLONE LENGTH: This sequence represents the entire insert of this
            clone unless otherwise noted. If there are overlapping clones, the
            overlaps are noted in the beginning and end of the Features
            listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550).
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
```

Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double strandedsequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation asLow Coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated average error rate is less than 1 per 10,000 bases using the Consed quality parameters. Regions that do not meet this requirement are annotated as Low Quality.

```
-----------------Summary Statistics
Center project name: ALH
Sequencing vector: pSMART; AF399742
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap version 0.990319
Contig length: 115035
Fraction of Phrap value < 40: 0
Error Rate in Consed: 0.00 per 10,000 bases
Number of N's in consensus: 0

------------ Distribution of Quality < 40 Bases:

          1000|
           900|
           800|
   #       700|
  bases    600|
           500|
           400|
           300|
           200|
           100|
             0|
               ----------------------------------------
                 5    10   15   20   25   30   35   40
                          Phrap Value Range
-----------------------.
```

```
FEATURES             Location/Qualifiers
     source          1. .115135
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="ES cell line"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /clone="MGS1-117K9"
     repeat_region   190. .274
                     /rpt_family="ID_B1"
     repeat_region   275. .497
                     /rpt_family="B3"
     repeat_region   498. .652
                     /rpt_family="ID_B1"
     repeat_region   760. .829
                     /rpt_family="L2"
     repeat_region   870. .990
                     /rpt_family="ORR1D"
     STS             1173. .1422
                     /standard_name="C79952"
     repeat_region   2880. .2914
                     /rpt_family="(TG)n"
     repeat_region   complement(2918. .3013)
                     /rpt_family="RSINE1"
     repeat_region   3119. .3173
                     /rpt_family="B3A"
     repeat_region   3179. .3214
                     /rpt_family="(TC)n"
     repeat_region   3214. .3254
                     /rpt_family="(CA)n"
     repeat_region   3991. .4035
                     /rpt_family="(TCCA)n"
     repeat_region   complement(4036. .4222)
                     /rpt_family="ID_B1"
     repeat_region   4286. .4445
                     /rpt_family="B1_MM"
     repeat_region   complement(4491. .4576)
                     /rpt_family="ID_B1"
     repeat_region   complement(4597. .4738)
                     /rpt_family="B1_MM"
     repeat_region   complement(4752. .4894)
                     /rpt_family="B1F"
     repeat_region   complement(5555. .5769)
                     /rpt_family="B3"
     repeat_region   complement(5904. .6096)
                     /rpt_family="B2_Mm1"
     repeat_region   complement(6120. .6248)
                     /rpt_family="B1_MM"
     repeat_region   6919. .6969
                     /rpt_family="MIR"
     repeat_region   7018. .7087
                     /rpt_family="ID3"
     repeat_region   7088. .7117
                     /rpt_family="(CA)n"
     repeat_region   7118. .7139
                     /rpt_family="ID_B1"
     repeat_region   7154. .7366
                     /rpt_family="ORR1B1"
     repeat_region   7511. .7756
                     /rpt_family="ORR1D"
     repeat_region   8060. .8095
                     /rpt_family="(TG)n"
     repeat_region   8138. .8160
                     /rpt_family="(CA)n"
     repeat_region   complement(8166. .8332)
                     /rpt_family="RSINE1"
     repeat_region   complement(9310. .9380)
                     /rpt_family="MER5A"
     repeat_region   complement(9398. .9498)
                     /rpt_family="MER5B"
     repeat_region   complement(9499. .9639)
                     /rpt_family="B1F"
     repeat_region   complement(9643. .9714)
                     /rpt_family="tRNA-Ala-GCY"
     repeat_region   complement(9756. .9947)
                     /rpt_family="B3"
     repeat_region   complement(9949. .10011)
                     /rpt_family="MER5B"
     repeat_region   10570. .10683
                     /rpt_family="ID_B1"
     repeat_region   complement(10813. .11021)
                     /rpt_family="B3"
     repeat_region   11321. .11431
                     /rpt_family="PB1D7"
     repeat_region   11437. .11495
                     /rpt_family="B3"
     repeat_region   complement(11496. .11892)
                     /rpt_family="MTA"
```

```
  Query Match             44.4%;  Score 2743.2;  DB 10;  Length 115135;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2818;  Conservative    0;  Mismatches   13;  Indels    7;  Gaps    6;

Qy       3341 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 3400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105112 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 105053

Qy       3401 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 3460
```

```
Db  105052 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 104993

Qy    3461 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 3520
Db  104992 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 104933

Qy    3521 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 3580
Db  104932 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 104873

Qy    3581 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 3640
Db  104872 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 104813

Qy    3641 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 3700
Db  104812 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 104753

Qy    3701 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 3760
Db  104752 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 104693

Qy    3761 CAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGA 3820
Db  104692 CAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGA 104633

Qy    3821 TATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACT 3880
Db  104632 TATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACT 104573

Qy    3881 CTCCCTCATATTCTGCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGAC 3940
Db  104572 CTCCCTCATATTCTGCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGAC 104513

Qy    3941 ACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTC 4000
Db  104512 ACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTC 104453

Qy    4001 CCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGAT 4060
Db  104452 CCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGAT 104393

Qy    4061 CAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGT 4120
Db  104392 CAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGT 104333

Qy    4121 GTCTCCAGTGCTGTGTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGC 4180
Db  104332 GTCTCCAGTGCTGTGTCTACAAAGACAGCGCCAGCCCCTCTTCTGTCAGCTGTGCTGAGC 104273

Qy    4181 AGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCAT 4240
Db  104272 AGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCAT 104213

Qy    4241 CCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTT 4300
Db  104212 CCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTT 104153

Qy    4301 GAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTC 4360
Db  104152 GAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTC 104093

Qy    4361 GCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCT 4420
Db  104092 GCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCT 104033

Qy    4421 TGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGT 4480
Db  104032 TGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGT 103973

Qy    4481 TTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAAC 4540
Db  103972 TTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAAC 103913

Qy    4541 CTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACC 4600
Db  103912 CTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACC 103853

Qy    4601 CCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGT 4660
Db  103852 CCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGA-CCCACTCCAAGT 103794

Qy    4661 AGATCTGGAAAAGACCCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGA 4720
Db  103793 AGATCTGGAAAAGA-CCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGA 103735

Qy    4721 AGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGG 4780
Db  103734 AGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGG 103675

Qy    4781 GAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTG 4840
Db  103674 GAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTG 103615

Qy    4841 TGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGG 4900
Db  103614 TGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGG 103555

Qy    4901 GGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAG 4960
Db  103554 GGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAG 103495

Qy    4961 AGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATT 5020
Db  103494 AGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATT 103435

Qy    5021 TTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTG 5080
Db  103434 TTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTG 103375

Qy    5081 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 5140
Db  103374 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 103315

Qy    5141 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 5200
Db  103314 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 103255

Qy    5201 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 5260
Db  103254 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 103195

Qy    5261 TTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTC 5320
Db  103194 TTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTC 103135

Qy    5321 TGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTC 5380
Db  103134 TGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTC 103075

Qy    5381 CTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCG 5440
Db  103074 CTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCG 103015

Qy    5441 TATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACG 5500
Db  103014 TATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACG 102955

Qy    5501 GCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCC 5560
Db  102954 GCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCC 102895

Qy    5561 TGGTTAGATTCTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTA 5620
Db  102894 TGGTTAGATTTTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTA 102835
```

```
Qy       5621 GCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGT 5680
              |||||||||||||||||||||||||||||||||||||||||||||||    | |   |||
Db     102834 GCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTCTGGT 102777

Qy       5681 TTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG- 5739
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102776 TTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGT 102717

Qy       5740 TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAA 5799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102716 TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAA 102657

Qy       5800 AGTAGCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTA 5859
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     102656 AGTAGCCAGAACTCACTAGCATTATATATGATTTGGGTTCTGAGTAACTGGGGAGTGTTA 102597

Qy       5860 GCTTTGTGACTTTGTAGCACCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCA 5919
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     102596 GCTTTGTGACTTTGTAGCAGCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCA 102537

Qy       5920 TTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATT 5979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102536 TTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATT 102477

Qy       5980 CCTTTTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG-G 6038
               |||||||||||||||||||||||||| ||||||||||||||||||||||||||||| |
Db     102476 TCTTTTCTCGAAGCATGACAAGTGTAT-GCAATACTTACATGCTCACTCGTTTACCTGTG 102418

Qy       6039 CTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAAT 6098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102417 CTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAAT 102358

Qy       6099 ATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTA 6158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102357 ATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTA 102298

Qy       6159 ATAAACTATTTTCCAGAC 6176
              ||||||||||||||||||
Db     102297 ATAAACTATTTTCCAGAC 102280
```

```
RESULT 10
AC110235/c
LOCUS       AC110235               158055 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-100M12, WORKING DRAFT SEQUENCE, 17 ordered
            pieces.
ACCESSION   AC110235
VERSION     AC110235.5  GI:20800205
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-100M12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 15, 2002 this sequence version replaced gi:20531900.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22383
                Center clone name: 100_M_12
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 153807 bases at least Q40
                Consensus quality: 155134 bases at least Q30
                Consensus quality: 155784 bases at least Q20
                Insert size: 218000; agarose-fp
                Insert size: 156455; sum-of-contigs
                Quality coverage: 7.1 in Q20 bases; agarose-fp
                Quality coverage: 10.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     6459: contig of 6459 bp in length
            *     6460     6559: gap of 100 bp
            *     6560     6667: contig of 108 bp in length
```

```
 *     6668     6767: gap of 100 bp
 *     6768     7863: contig of 1096 bp in length
 *     7864     7963: gap of 100 bp
 *     7964     9426: contig of 1463 bp in length
 *     9427     9526: gap of 100 bp
 *     9527    11011: contig of 1485 bp in length
 *    11012    11111: gap of 100 bp
 *    11112    12741: contig of 1630 bp in length
 *    12742    12841: gap of 100 bp
 *    12842    14789: contig of 1948 bp in length
 *    14790    14889: gap of 100 bp
 *    14890    17826: contig of 2937 bp in length
 *    17827    17926: gap of 100 bp
 *    17927    21244: contig of 3318 bp in length
 *    21245    21344: gap of 100 bp
 *    21345    27413: contig of 6069 bp in length
 *    27414    27513: gap of 100 bp
 *    27514    36381: contig of 8868 bp in length
 *    36382    36481: gap of 100 bp
 *    36482    45104: contig of 8623 bp in length
 *    45105    45204: gap of 100 bp
 *    45205    58369: contig of 13165 bp in length
 *    58370    58469: gap of 100 bp
 *    58470    85782: contig of 27313 bp in length
 *    85783    85882: gap of 100 bp
 *    85883   112706: contig of 26824 bp in length
 *   112707   112806: gap of 100 bp
 *   112807   153927: contig of 41121 bp in length
 *   153928   154027: gap of 100 bp
 *   154028   158055: contig of 4028 bp in length.
FEATURES             Location/Qualifiers
     source          1..158055
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-100M12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1..6459
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    6560..6667
                     /note="assembly_fragment"
     misc_feature    6768..7863
                     /note="assembly_fragment"
     misc_feature    7964..9426
                     /note="assembly_fragment"
     misc_feature    9527..11011
                     /note="assembly_fragment"
     misc_feature    11112..12741
                     /note="assembly_fragment"
     misc_feature    12842..14789
                     /note="assembly_fragment"
     misc_feature    14890..17826
                     /note="assembly_fragment"
     misc_feature    17927..21244
                     /note="assembly_fragment"
     misc_feature    21345..27413
                     /note="assembly_fragment"
     misc_feature    27514..36381
                     /note="assembly_fragment"
     misc_feature    36482..45104
                     /note="assembly_fragment"
     misc_feature    45205..58369
                     /note="assembly_fragment"
     misc_feature    58470..85782
                     /note="assembly_fragment"
     misc_feature    85883..112706
                     /note="assembly_fragment"
     misc_feature    112807..153927
                     /note="assembly_fragment"
     misc_feature    154028..158055
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    41469 a  36830 c  36521 g  41635 t   1600 others
ORIGIN

  Query Match             44.4%;  Score 2743.2;  DB 2;  Length 158055;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2818;  Conservative    0;  Mismatches   13;  Indels    7;  Gaps    6;

Qy       3341 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 3400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73684 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 73625

Qy       3401 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 3460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73624 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 73565

Qy       3461 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 3520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73564 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 73505

Qy       3521 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 3580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73504 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 73445

Qy       3581 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 3640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73444 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 73385

Qy       3641 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 3700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73384 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 73325

Qy       3701 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 3760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73324 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 73265

Qy       3761 CAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGA 3820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73264 CAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGA 73205

Qy       3821 TATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACT 3880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73204 TATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACT 73145

Qy       3881 CTCCCTCATATTCTGCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGAC 3940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73144 CTCCCTCATATTCTGCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGAC 73085

Qy       3941 ACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTC 4000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73084 ACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTC 73025

Qy       4001 CCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGAT 4060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73024 CCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGAT 72965

Qy       4061 CAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGT 4120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72964 CAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGT 72905

Qy       4121 GTCTCCAGTGCTGTGTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGC 4180
              ||||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db      72904 GTCTCCAGTGCTGTGTCTACAAAGACAGCGCCAGCCCCTCTTCTGTCAGCTGTGCTGAGC 72845

Qy       4181 AGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCAT 4240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72844 AGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCAT 72785

Qy       4241 CCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTT 4300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72784 CCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTT 72725
```

```
Qy       4301 GAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTC 4360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72724 GAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTC 72665

Qy       4361 GCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCT 4420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72664 GCACCCTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCT 72605

Qy       4421 TGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGT 4480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72604 TGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGT 72545

Qy       4481 TTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAAC 4540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72544 TTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAAC 72485

Qy       4541 CTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACC 4600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72484 CTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACC 72425

Qy       4601 CCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGT 4660
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      72424 CCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGA-CCCACTCCAAGT 72366

Qy       4661 AGATCTGGAAAAGACCCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGA 4720
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      72365 AGATCTGGAAAAGA-CCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGA 72307

Qy       4721 AGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGG 4780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72306 AGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGG 72247

Qy       4781 GAGTGAAACCAAGAGGCTGAAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTG 4840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72246 GAGTGAAACCAAGAGGCTGAAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTG 72187

Qy       4841 TGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGG 4900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72186 TGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGG 72127

Qy       4901 GGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAG 4960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72126 GGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAG 72067

Qy       4961 AGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATT 5020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72066 AGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATT 72007

Qy       5021 TTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTG 5080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72006 TTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTG 71947

Qy       5081 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 5140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71946 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 71887

Qy       5141 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 5200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71886 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 71827

Qy       5201 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 5260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71826 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 71767

Qy       5261 TTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTC 5320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71766 TTTCCCTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTC 71707

Qy       5321 TGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTC 5380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71706 TGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTC 71647

Qy       5381 CTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCG 5440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71646 CTTCTCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCG 71587

Qy       5441 TATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACG 5500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71586 TATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACG 71527

Qy       5501 GCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCC 5560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71526 GCCAACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCC 71467

Qy       5561 TGGTTAGATTCTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTA 5620
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db      71466 TGGTTAGATTTTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTA 71407

Qy       5621 GCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGT 5680
              |||||||||||||||||||||||||||||||||||||||||||||||    | |   |||
Db      71406 GCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTCTGGT 71349

Qy       5681 TTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG- 5739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db      71348 TTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGT 71289

Qy       5740 TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAA 5799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71288 TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAA 71229

Qy       5800 AGTAGCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTA 5859
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      71228 AGTAGCCAGAACTCACTAGCATTATATATGATTTGGGTTCTGAGTAACTGGGGAGTGTTA 71169

Qy       5860 GCTTTGTGACTTTGTAGCACCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCA 5919
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      71168 GCTTTGTGACTTTGTAGCAGCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCA 71109

Qy       5920 TTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATT 5979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71108 TTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATT 71049

Qy       5980 CCTTTTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG-G 6038
               |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| |
Db      71048 TCTTTTCTCGAAGCATGACAAGTGTAT-GCAATACTTACATGCTCACTCGTTTACCTGTG 70990

Qy       6039 CTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAAT 6098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70989 CTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAAT 70930

Qy       6099 ATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTA 6158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70929 ATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTA 70870

Qy       6159 ATAAACTATTTTCCAGAC 6176
              ||||||||||||||||||
Db      70869 ATAAACTATTTTCCAGAC 70852
```

```
RESULT 11
AB046848
LOCUS       AB046848                5549 bp    mRNA    linear   PRI 22-FEB-2001
DEFINITION  Homo sapiens mRNA for KIAA1628 protein, partial cds.
ACCESSION   AB046848
VERSION     AB046848.1  GI:10047332
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Kikuno,R., Nakayama,M., Hirosawa,M. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XVIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 7 (4), 273-281 (2000)
```

```
  MEDLINE   20450683
   PUBMED   10997877
REFERENCE   2  (bases 1 to 5549)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..5549
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="fh14247"
                     /note="vector:pBluescriptII SK plus"
     gene            1..5549
                     /gene="KIAA1628"
     CDS             <81..3023
                     /gene="KIAA1628"
                     /note="Start codon is not identified. fh14247 cDNA clone
                     for KIAA1628 has a 40-bp different sequence replacing a
                     108-bp sequence at the position between 737 and 846, and
                     an 88-bp insertion after the position 983 of the sequence
                     of KIAA1628, respectively."
                     /codon_start=1
                     /product="KIAA1628 protein"
                     /protein_id="BAB13454.1"
                     /db_xref="GI:10047333"
                     /translation="MGRRKGRSALSRSSLERVNQGEVKVIGRGAREGGGGQAWGTGGG
                     GRDLSCGIPRSASPLAPLAAPAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNG
                     APLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPS
                     APTRVTATPLSSSAVLVAWERPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTEL
                     QVRDLEPNTDYEFYVVAYSQLGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVA
                     WLPLPPSLSNGQVVKYKIEYGLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAG
                     TAAGFGAPSQWMHHRTPSMHNQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYK
                     LYWREVGAEEEANGDRLPGGRGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAF
                     NKHEDGYAAVWKGKTEKAPAPDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTV
                     KIVNYTVRFSPWGLRNASLVTYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGS
                     VVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEH
                     QWTLLTTQGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLD
                     MHSVTGIIVGVCLGLLCLLACMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPP
                     SPPAAHELESLVHPHPQDWSPPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSG
                     LSWAGSWAGCELPQAGPRPALTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACR
                     NQVEAEVIVHSDFSASNGNPDLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDREL
                     GGCELAAPGPDRLTCLPEAASASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLP
                     RSPVSSSA"
BASE COUNT     1157 a   1546 c   1496 g   1350 t
ORIGIN

  Query Match             43.6%;  Score 2691;  DB 9;  Length 5549;
  Best Local Similarity   74.2%;  Pred. No. 0;
  Matches 3968;  Conservative    0;  Mismatches 1155;  Indels  224;  Gaps   36;

Qy        991 CTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACC 1050
Db        264 CTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACA 323

Qy       1051 GCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGAC 1110
Db        324 GCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAAC 383

Qy       1111 GGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTC 1170
Db        384 GGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTC 443

Qy       1171 ATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCG 1230
Db        444 ATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCG 503

Qy       1231 GGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCC 1290
Db        504 GGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCC 563

Qy       1291 CCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGG 1350
Db        564 CCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGG 623

Qy       1351 CCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGA 1410
Db        624 CCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGC 683

Qy       1411 GTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGG 1470
Db        684 ATGGACAATGTGGAATACCAGTTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGG 743

Qy       1471 GACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCC 1530
Db        744 GACCTGGAACCCAACACAGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCC 803

Qy       1531 AGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCC 1590
Db        804 AGCCGCACCTCCACCCCAGCACTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCC 863

Qy       1591 CAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCC 1650
Db        864 CAGCTCTCCCTGTCCAGCCCCAACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCC 923

Qy       1651 TCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAA 1710
Db        924 CCCAGCCTGAGCAATGGGCAGGTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAA 983

Qy       1711 GATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTT 1770
Db        984 GATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTT 1043

Qy       1771 CAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTC 1830
Db       1044 CAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCC 1103

Qy       1831 CCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTT 1890
Db       1104 CCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTT 1163

Qy       1891 GCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCG 1950
Db       1164 GCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCA 1223

Qy       1951 CCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAG 2010
Db       1224 CCCCCTCACCCCACCCAGATCTCTGGCTACAAACTATATTGGCGGAGGTGGGGGCTGAG 1283

Qy       2011 GAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGG 2070
Db       1284 GAGGAGGCCAATGGCGATCGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGG 1343

Qy       2071 CCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGG 2130
Db       1344 CCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGG 1403

Qy       2131 CCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGG 2190
Db       1404 CTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGG 1463

Qy       2191 AAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTG 2250
Db       1464 AAGGGCAAGACGGAGAAGGCGCCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTG 1523

Qy       2251 CCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAG 2310
Db       1524 CCTCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAA 1583

Qy       2311 AAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGG 2370
Db       1584 AAGCCAGATTTCACCACAGTCAAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGG 1643
```

```
Qy  2371 CTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGC 2430
Db  1644 CTCAGGAATGCCTCCCTGGTCACCTATTACACCAGTTCTGGAGAAGACATCCTCATTGGC 1703

Qy  2431 GGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGAT 2490
Db  1704 GGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGAT 1763

Qy  2491 GGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCT 2550
Db  1764 GGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCA 1823

Qy  2551 TCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCC 2610
Db  1824 TCCGACCTGCGACTGAGCCCCCTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCC 1883

Qy  2611 ACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAG 2670
Db  1884 ACAGAGCCCAACGGGGAGATCGTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAG 1943

Qy  2671 CCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCAT 2730
Db  1944 CCTGAGCACCAGTGGACCTTGCTCACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCAT 2003

Qy  2731 GGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCT 2790
Db  2004 GGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCT 2063

Qy  2791 GGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTG 2850
Db  2064 GGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTG 2123

Qy  2851 GATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTG 2910
Db  2124 GACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTG 2183

Qy  2911 GCCTGCATGTGTGCTGGCCTACGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCC 2970
Db  2184 GCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCC 2243

Qy  2971 TCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGT 3030
Db  2244 TCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGC 2303

Qy  3031 GTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCA 3090
Db  2304 CCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCG 2363

Qy  3091 CCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCA 3150
Db  2364 CCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCT 2423

Qy  3151 GATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGG 3210
Db  2424 GAAGGCCGGAGTCACTCCAAAAGAAAGATCTCCTGGGCTCAACCAAGCGGGCTGAGCTGG 2483

Qy  3211 GCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACC 3270
Db  2484 GCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAGG---CAGGCCCCCGGCCGGCTCTGACC 2540

Qy  3271 CGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTAT 3330
Db  2541 CGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGCTCTGGTGTAC 2600

Qy  3331 GACGGCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAA 3390
Db  2601 GACGCCATAAAGGGCAATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAG 2660

Qy  3391 GCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTC 3450
Db  2661 GCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTC 2720

Qy  3451 CAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCT 3510
Db  2721 CAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTT 2780

Qy  3511 GTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACA 3570
Db  2781 GGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCA 2840

Qy  3571 ACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGC 3630
Db  2841 GCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTAC 2900

Qy  3631 TCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAA 3690
Db  2901 CCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAA 2960

Qy  3691 GCCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAG 3750
Db  2961 TCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCTGCC 3020

Qy  3751 GTCCCCTGAGCAGA----------------AGGCAGATATGGCTCAGGAACAT------- 3787
Db  3021 TAGCTCTTCCCAGAGGATGTGGTTTGGGGCAGGCAGGTATGGATCACATAGGATGCGATA 3080

Qy  3788 ------GCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGCC 3841
Db  3081 CCTGTGGCCGTGTATGTCCACATGTGTGCCTGTAGATACATCATCAAGCCCTTTGGAGCT 3140

Qy  3842 TCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAA------CTTTACTCTCCCTCATATTCTG 3895
Db  3141 TCCTAAGTTGCTTTGGCTGAGGGGAGAGGAAAACATGGATTATTCACTCCCCCCATACTC 3200

Qy  3896 CATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTG 3955
Db  3201 TTTGTGATACACATGTGACATGTGAAAGACATACGAGACATAGCTACATGTGATGTGCAC 3260

Qy  3956 TGTGTGTGTGTGTGTGTGTG-----TGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGC 4010
Db  3261 ATGTGTGAAGTGCATGTATGCGTACTGGTTGTTGAGCTGGGAAACCGTGGCCCAGGCAGT 3320

Qy  4011 ACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTG--TAACAGTGATCAGT-GCC 4067
Db  3321 GGTCACTACAGCCTGATTGGTCCTCCAGGTCAGAACGGTGCCCCACAGTGGTCAGTCCCC 3380

Qy  4068 AGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATT--ACACTCTGAGAGTGTCTC 4125
Db  3381 AGCCCTGTGGGCCCCCACCTCCATCGCCCAGCCTTTTATTACACACTCTGAGAGTGTCTC 3440

Qy  4126 CAGTGCTGTGTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGT 4185
Db  3441 CAATGCCTGTCTGACAAAGACAGTCCCAGCCCATTCTC---CTGTCTGGCTGGGTTGGGT 3497

Qy  4186 GCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCC 4245
Db  3498 GCAA------------------------GCAGGCTCTGAATGCCTGGCATTTCAGCTGC 3532

Qy  4246 TGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCA 4305
Db  3533 ATCACCTCCCAGCTCCTTATTGCCCAAATAGAGAGGGTGGCCCTGGCTCCCCTCCGAGCA 3592

Qy  4306 TCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACC 4365
Db  3593 ACTCTGCATTTAATTTTGTAATCTGGGAAGTGCCTGGTTTTGAAAATCCGCTTTCTCTCA 3652

Qy  4366 CTGTCCCTCTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCT-CATGTCATCTTGCT 4424
Db  3653 CTCTTCCCCTCCTTCCTTGCCCCTGGCTGCTCTAGTGTTCTGTCTCCCAGTCACCTCGCT 3712

Qy  4425 CGATTATC-CTGGGGCCCTTCTCTT--TCCCATGATGCCCCTGATTTCCTCACTGCTGTT 4481
Db  3713 CTCCCAGCACCAGTGCCCTTCTCCTGCTCCCAGATACTCTTTCCTTTCCTCTCTCCTGTT 3772

Qy  4482 TTCATT---------TCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTG 4532
```

```
Db   3773 TTCCTTCCTCTGCTATCTCTCACACCTCTCCCAGACTATGTCATCTTGTTCTCCTGCCTG 3832
Qy   4533 AGTTCAACCTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCA------- 4585
Db   3833 GGTTCAAACTCTGCATCCTTCTCTAACAACGTGACTACCTCATGTCTGCTTCAAGGCCCC 3892
Qy   4586 --GACCATAGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCT------ 4637
Db   3893 CGTGCCCTTCCTGTATCCGCGGCTGCCGCGCACTCGCCTGCCATCCTCCTGCCTCCTCTT 3952
Qy   4638 --------------------------GGGCAGATGAGCCCACTCCAAGTAGATCTGGA 4669
Db   3953 CACTCAGTGCTTCTGCTTGCCCTGCCCCAGGCAGCCCACCCACGCCCAGTGCGGGTGTGG 4012
Qy   4670 AAAGACCCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCT 4729
Db   4013 AGAAGATCTTCTGGCTTCCCTGCAT--CTTGCCTTTGGGATTGGGATCCAAGGGTTCTCC 4070
Qy   4730 ATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAAC 4789
Db   4071 ATGGATG----GATCCAAGTCATAGAGGGGAATGTTTGAGACAGGGAAGGGGGCTGTGAT 4126
Qy   4790 CAAGAGGCTGAAAAAAAATGGCTGCCACCC---ATCTGCACAGAGAGATGGGTGTGTGCT 4846
Db   4127 CCAGAGGCTCAGAATAAAAAGATGCCCTCCCTTCTATGCAGGGGGGCAAGTTTACTGGAT 4186
Qy   4847 TTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGC 4906
Db   4187 GGAGATGATTTGGGCCTCTCTTCCAGAAGAAGCTAAAGGAAGAGAAGGGGAGTGAGAGTT 4246
Qy   4907 CATGGCTCAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTT- 4965
Db   4247 CAGGGAGGCCCTTCCCACCCTGTGAGGCTTGACTTGATCTGGATTGGGGATGACAGGAAT 4306
Qy   4966 -----CCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATT 5020
Db   4307 CTCACCCTCTGGGGTGCTGGCAAGGAGGTCTTTGCACAGGAAAAGG--GGTAGCTCATTT 4364
Qy   5021 TTATTTTATCTTTCTTTACACTGAATCCCAAAATCAT----------CTTACCACAAAGG 5070
Db   4365 CAGTTTGTTTTTCTTTAAATTGAATCCTCAAGTCATTTTCTGTTCACCTGCCGCACAGG 4424
Qy   5071 GCCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCAC 5130
Db   4425 GACAAGCTTGACTTCTATTTTCTGTGTAGTGAAAACAATGTCATTTATTTGGTTTTTCAC 4484
Qy   5131 CTCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTT 5190
Db   4485 CTCAGCCCTCTCATAGGAGCATAGAATGTTAGGGTCTTTACTCCCTA------------- 4531
Qy   5191 CTATTTCAACTGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATA 5250
Db   4532 --ATGATGTCTGATTGGCACATCAAGAGTTAACTCTGCCTTCTGGGCCAAATTCGAAATA 4589
Qy   5251 ACCAGTCTATTTTCCCTTTATTTTATTTTATTTTATTTT------ATTTTATGTCTTTCA 5304
Db   4590 ACCAGTCCATTTTTCCTTTTTTTTTTTTTTTTTTTTTTTAAATGGTGGAATGTCTCTCA 4649
Qy   5305 GTGGAGTTGTAGCTT---------CTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCAC 5355
Db   4650 GCACAGTTGCGGCTTCCTCAAACCCTGAAAGCATCTGTGTTTATTATACTCGGGTGTCAC 4709
Qy   5356 TCATGTTTGA---CCCCACCCACATTTCCTTCTCCTCCCC--TCTTCAGCCAGCCTATGA 5410
Db   4710 TCACTGTTGATGTCTGCACCTACGTTTCCACCTCCTCCCCCTCCTTCAGCCAGCCTATGA 4769
Qy   5411 TAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGA 5470
Db   4770 TAACACTAAAGATTATTAATGTTGGTTTTGTATCTCGTTAAAGACAGAATTGTCACTTGT 4829
Qy   5471 ACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCAC--CGTATGTTTCTTCATTG 5528
Db   4830 AGTATTTCTGTAGCATTCAGCGCTGCTGTGGCTAACACCACTGTGTATGTTTCATCATTG 4889
Qy   5529 CTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCCT------- 5581
Db   4890 CTCTGAAGGTCAAAAGCCTCATTTTATTTTGCTGGTTTGATTTTTTTTTTTAAAGAAGA 4949
Qy   5582 ---------TGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCAC 5632
Db   4950 AAAAAAAACTGCCCTGAATTAAATGGCTGTTTTAACAGTAGGCTCTTAGCATTATACCAC 5009
Qy   5633 ATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATA 5692
Db   5010 ATAGTCATTTTTCATGTTCTTGTTTAACAGGCACT--GAGGTTCTGGTTTAAATTAAATA 5067
Qy   5693 GCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCA 5751
Db   5068 GCTGCAAATGAGACAATTTATAACCCATTAGGTTGGGTGGAAAATTGTTTCTCAAAAGCA 5127
Qy   5752 AATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAAC 5811
Db   5128 AATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTGGCCATTTC 5187
Qy   5812 TCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTT 5871
Db   5188 TCACTAGCACTATATATGATTTGGGCTCTGGGTAATTTGGAAGTGTTAGGTTTGTGTCTT 5247
Qy   5872 TGTAGCACCAGGTCT-TATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTT 5930
Db   5248 TGTAGCAGTATTTTTATTAGAAAAGAATCTATTGGCCTTTTACAGGGTATTAATCCCTTT 5307
Qy   5931 GTCGTTTGCCATGGATGCCTTAAGTTCTTTG-GAGTCTCATTTAAGAATTCCTTTTCTCG 5989
Db   5308 GTCACCTACCATTGATGCCTTAAGTTTTCTGAGTCTCAATTAAAAATCTTCCTTTTCTTG 5367
Qy   5990 AAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGT 6048
Db   5368 ATGCATGACAAGTGT-----AATCAGTACTTGCTCATTTATTTGTCTGTATTTAGTTTAT 5422
Qy   6049 GCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCT 6108
Db   5423 GCTGTACTATTTAATTATCCTTCCAGCGTTTTTTTTTTCTCCTTACAAATATGATACTCT 5482
Qy   6109 TTATTGTTACACTAAGGTGTTGA-TCATGTATCTGTCCCTGTAAAGAATTAATAAACTAT 6167
Db   5483 TTAGTGTTAAGCTAAGGCATTGATTCATGTATCTGTCCTTATAATGAATTAATAAACTAT 5542
Qy   6168 TTTCCAG 6174
Db   5543 TTTCCAG 5549
```

RESULT 12
AX418406

```
LOCUS       AX418406                3741 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0206329.
ACCESSION   AX418406
VERSION     AX418406.1  GI:21523336
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Rastelli,L., Shimkets,R.A., Zerhusen,B., Malyankar,U.M. and
            Padigaru,M.
  TITLE     Human polynucleotides and polypeptides encoded thereby
  JOURNAL   Patent: WO 0206329-A 1 24-JAN-2002;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3741
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
                     /note="NOVX nucleic acid with homology to NOPE-like
```

```
                         protein"
BASE COUNT      708 a    1209 c    1150 g    673 t      1 others
ORIGIN

  Query Match             43.4%;  Score 2681.8;  DB 6;  Length 3741;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 3141;  Conservative    0;  Mismatches  583;  Indels   30;  Gaps    8;

Qy          3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62
Db          9 GGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 68

Qy         63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122
Db         69 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACT 128

Qy        123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182
Db        129 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 188

Qy        183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242
Db        189 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 248

Qy        243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302
Db        249 CCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 308

Qy        303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362
Db        309 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCT 365

Qy        363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422
Db        366 AGCCCACGGNCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCAGTCT 425

Qy        423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482
Db        426 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 485

Qy        483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542
Db        486 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 545

Qy        543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602
Db        546 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 605

Qy        603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662
Db        606 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 665

Qy        663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722
Db        666 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACCAGGGGGCA 725

Qy        723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782
Db        726 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 785

Qy        783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842
Db        786 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA---GACGG 842

Qy        843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902
Db        843 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 902

Qy        903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962
Db        903 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 962

Qy        963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTT---GCTGCCCCAGCCATCTCGCAGGCACC 1019
Db        963 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACTCAGGCGCC 1022

Qy       1020 CGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGA 1079
Db       1023 CGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGA 1082

Qy       1080 GCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGT 1139
Db       1083 GCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGT 1142

Qy       1140 CAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGG 1199
Db       1143 CAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGG 1202

Qy       1200 CTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGC 1259
Db       1203 CTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGC 1262

Qy       1260 GGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAG 1319
Db       1263 CGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAG 1322

Qy       1320 CAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGG 1379
Db       1323 CAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGG 1382

Qy       1380 CTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGT 1439
Db       1383 CTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAGTTTGCAGT 1442

Qy       1440 AAACAATGACACCACAGAGCTGCAGGTTCGGACCTGGAACCCAACACGGATTATGAGTT 1499
Db       1443 GAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTT 1502

Qy       1500 CTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCA 1559
Db       1503 CTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCA 1562

Qy       1560 TACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTC 1619
Db       1563 CACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTC 1622

Qy       1620 GGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAA 1679
Db       1623 GGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAA 1682

Qy       1680 GTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGG 1739
Db       1683 GTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGG 1742

Qy       1740 AAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGAT 1799
Db       1743 AAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGAT 1802

Qy       1800 TTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACC 1859
Db       1803 TTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCC 1862

Qy       1860 TGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAA 1919
Db       1863 CAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAA 1922

Qy       1920 GATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATA 1979
Db       1923 GATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTA 1982

Qy       1980 CAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGG 2039
Db       1983 CAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGG 2042

Qy       2040 GGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCA 2099
```

```
Db   2043 GGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCA 2102

Qy   2100 GTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAA 2159
Db   2103 GTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAA 2162

Qy   2160 CAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCC 2219
Db   2163 CAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACC 2222

Qy   2220 AGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAA 2279
Db   2223 AGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAA 2282

Qy   2280 CAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGT 2339
Db   2283 CAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGT 2342

Qy   2340 CAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTA 2399
Db   2343 CAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTA 2402

Qy   2400 TACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTT 2459
Db   2403 ---CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTT 2459

Qy   2460 TGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTC 2519
Db   2460 TGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTC 2519

Qy   2520 CACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACC 2579
Db   2520 CACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACC 2579

Qy   2580 ATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTA 2639
Db   2580 GTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTA 2639

Qy   2640 TCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCAC 2699
Db   2640 TCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCAC 2699

Qy   2700 ------AGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2700 GCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTA 2759

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2760 CTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGA 2819

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCAT 2873
Db   2820 TGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCAT 2879

Qy   2874 CATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACG 2933
Db   2880 CATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCG 2939

Qy   2934 ACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCC 2993
Db   2940 CCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCC 2999

Qy   2994 AGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGA 3053
Db   3000 CGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGA 3059

Qy   3054 GTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAA 3113
Db   3060 GTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAG 3119

Qy   3114 GGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAG 3173
Db   3120 GGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAG 3179

Qy   3174 AAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGA 3233
Db   3180 AAAGGT---AAGTGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGA 3236

Qy   3234 GCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGG 3293
Db   3237 GCTGCCCCAGG---CAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGG 3293

Qy   3294 AACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAG 3353
Db   3294 AACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGATC------AGGGCAATGGGAG 3347

Qy   3354 AAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGA 3413
Db   3348 GAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGA 3407

Qy   3414 CTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACC 3473
Db   3408 CTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCC 3467

Qy   3474 ACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGA 3533
Db   3468 CCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGC 3527

Qy   3534 CTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCAC 3593
Db   3528 CTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTAC 3587

Qy   3594 CTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGC 3653
Db   3588 CTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGT 3647

Qy   3654 TATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAG 3713
Db   3648 GCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAG 3707

Qy   3714 CCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCT 3747
Db   3708 CCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741
```

RESULT 13
AC106509/c

```
LOCUS       AC106509               224227 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-202I8, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC106509
VERSION     AC106509.4  GI:30580817
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 224227)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
```

```
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 224227)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:22857495.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLBA
                Center clone name: CH230-202I8
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 211258 bases at least Q40
                Consensus quality: 214044 bases at least Q30
                Consensus quality: 215992 bases at least Q20
                Estimated insert size: 220399; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1  221791: contig of 221791 bp in length
            *     221792  221891: gap of unknown length
            *     221892  222909: contig of 1018 bp in length
            *     222910  223009: gap of unknown length
            *     223010  224227: contig of 1218 bp in length.
FEATURES             Location/Qualifiers
     source          1..224227
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-202I8"
     misc_feature    1..1752
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3617..3835
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH334357"
     misc_feature    220120..221791
                     /note="wgs_contig"
BASE COUNT    58284 a  52300 c  50849 g  55959 t   6835 others
ORIGIN

  Query Match             29.8%;  Score 1843.2;  DB 2;  Length 224227;
  Best Local Similarity   84.5%;  Pred. No. 0;
  Matches 2456;  Conservative    0;  Mismatches  293;  Indels  157;  Gaps   28;

Qy       3341 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 3400
              ||||||||||||||||||||||||||||||||||||||||||||||||| || || ||||
Db      98567 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAGGCCGAAGTCA 98508

Qy       3401 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 3460
              ||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db      98507 TTGTCCACTCCGATTTCAGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 98448

Qy       3461 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 3520
              |||| |||||  | ||| ||||||||  ||||||| ||||  |||| |  |||||| |||
Db      98447 AGCCCGAGGAGGCCCTGCCTGCAGAGGTTCTGCCTCCCACTGCTGGGGTCGTGGATTTGT 98388

Qy       3521 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 3580
              | ||||||||||||||||||||||||||||||| || |||||||||| |||||| |||||
Db      98387 CCCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGAAGAGTGCCAACCAGCAACCACTGGGC 98328

Qy       3581 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 3640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98327 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 98268

Qy       3641 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 3700
              ||||||||| |||| |||||||||||||||||||| ||||||||||||||||||||||||
Db      98267 AGCCCAGCAGTGCTGTAGAGGAGGCCCCTGGGAAAGGCTGCCAGCCCAAAGCCCTGTGTC 98208

Qy       3701 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 3760
              |||||  | ||||||||||||||||||| ||||||||||||||||||||||| |||||||
Db      98207 CTCTAGAAATCAGCCCAAGCCTTCCCAGAGCCCCTGTCTCCTCTGCTCAGGTGCCCTGAG 98148

Qy       3761 CAGAAGGCAGATATGGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGA 3820
              |||||||| |||| |||||||| ||||| |||||||||| ||||    |||||| |||||
Db      98147 CAGAAGGCTGATACGGCTCAGGGACATGACATGCATGGCCACAC---GGTGTACCAGAGA 98091

Qy       3821 TATCCATAAGTCCTTGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTA-- 3878
```

```
Db     98090 TATCAACAAGCCCTTTGGAGCCTCTAGGATCCTTTGGCTGGGGTGAGGGAGAACTTTATT 98031

Qy      3879 CTCTCCCTCATATTCTGCATCAC-------ATACAGGAGGGACTTGAGACACAGCTCTGT 3931
Db     98030 CTCTCCCCTATATTTTGCATCACATATGTGATACAGGAGGGACTTGAGACACAGCTGCAT 97971

Qy      3932 GTAATGGACACGTGTGAAGTCGTGTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGG 3991
Db     97970 GTGATGGACA--------------------------GGTGTGTGTGCTGGTTGAGCTGGG 97937

Qy      3992 AAACCTCTCCCTATGTAGCACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGT 4051
Db     97936 AAACCTCTCCCTATGTAGCAGTCACTGGGGTCTGATTGACCCTCCATGGCAGGACGGTGT 97877

Qy      4052 AACAGTGATCAGTGCCAGCTCTTTGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACAC 4111
Db     97876 ACCAGTGATCAGTGCCAGGTCTTTGAGC-TTTAGCCTTGTCACCTAGCCTTTTATTACAC 97818

Qy      4112 TCTGAGAGTGTCTCCAGTGCTGTGTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCT 4171
Db     97817 TCTGAGAGTGTCTCCAGTGCTGTGTCTACAAAGA-GGTGCCAGCCCCTCTTCGGTCAGCT 97759

Qy      4172 GTGCTGAGCAGAGTGCCAGTCAACTCCACGGGCCTATGACACCGCAGCCTACCACAGCAT 4231
Db     97758 GTGCTGAGCAGAGTGCCAGTCAA--CCGCAGGCCTACGACACGGCA-------ACAGCAC 97708

Qy      4232 GGCTGTCATCCCCCTGGCCTCCTAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCC 4291
Db     97707 GGCCGTCATCCCCCTGGCCTCCTAAGGTCCAGACGGCTGGGTGAACCCAG---------- 97658

Qy      4292 CTCTCCTTTGAGCATCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAA 4351
Db     97657 CTCTCCTTTGAGCAGCTCTGTACCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAA 97598

Qy      4352 TCTTCTTTCGCACCCTGTCCCTC-------TCTGCCCCTTCCTTCATTTGTTCTGGTGAT 4404
Db     97597 TCTTCTTTCGCCCCACCCCCATCCCCCAGTCCCTCGCTGCCCCTTCATTGTTCTGGTGGT 97538

Qy      4405 CTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTG 4464
Db     97537 CTGTCTCATGTCATCTTGCTCCCTAATCCTGGGGCCCTTCTCTTTCCCATGATGCCCCTT 97478

Qy      4465 ATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCT 4524
Db     97477 ATTTCCTCACGGCTGTTTCCATTTCTGTCTCCAACGCTTGTCTTTGTGTCGTGTGTTTCT 97418

Qy      4525 CGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTC 4584
Db     97417 CGTCCCTGAGTTCAACCTATGCACCCTTCCCTAACAACATGACTACCTCATGTCTGCTTC 97358

Qy      4585 AGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTGGGCAGA 4644
Db     97357 AGACCATAGTGTGACCGCTGGGTCCCCACAGCTCCCCTGCCAACTGCCTTCCTGGGCAGA 97298

Qy      4645 TGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTGTGGCTTGTCTGGCTGCCCTCCCCT 4704
Db     97297 TGA-TCCACTCCAGGTAGATCTGGAGAAGA-CTTTGTAGCTTCCCTGGCTGCCGTCTCCT 97240

Qy      4705 TGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGGCTGCACAGGGGAACC 4764
Db     97239 TGGTATT----------------------GAGATGAACCCAGGCCACACCAGGGAAGC 97204

Qy      4765 CCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATGGCTGCCACCCATCTG 4824
Db     97203 CCTGAGAAGGGGTAGGCACTGATACCAAAAGGCT-TAATAAAATGACTGCCCCCCATCTG 97145

Qy      4825 CACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAAACTGAAGGGGTGAGG 4884
Db     97144 CACAGAGAAATGAGTGCATGC-TTTGCCATGCAGTCCTGGCCAAAGCTGAAGGGGTGAGG 97086

Qy      4885 AGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTGGAGCCCTGAAC---- 4940

Db     97085 AGAAGGTAGTTACCAGAGCTGCAGTGACTCAGCTCCCTGACCTTGGAGCCTTGAACCTGA 97026

Qy      4941 CTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGTAAGGGAAGTCTTTGCATAGG 5000
Db     97025 CTGGATTCAGGGTGGCAAAGAGTTTCCTCCAAGATGCTAGAAGGGAAGTCTCTGCACAGG 96966

Qy      5001 AAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACACTGAATCCCAAAATCATCTT 5060
Db     96965 AAAAGGGCGGCTGGCTCATTTTGTTTTATCTTCCTTTACACTGAATCCAGAAATCATCTT 96906

Qy      5061 ACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCT 5120
Db     96905 ACCACATAGGGCCAAGCCTGATTGGTATTTCCTGAGTCACGGAAGCCATACCATCTCCTT 96846

Qy      5121 GGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCT 5180
Db     96845 GATTTCTCACTTCAGCCGTGTTCCAGGATCATC-GAATGTGGGATCTGTGCTCTCTTTCT 96787

Qy      5181 GCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAA 5240
Db     96786 GCACATCTTTCTATTT-TACTGGCTGGCACACCAATCATTAACTCTGGCTTCTGGGCCAA 96728

Qy      5241 GTTAGAAATAACCAGTCTATTTTCC----------------------------------- 5265
Db     96727 ATTAGAAATAACCAGTCTATCTCTCTATATCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 96668

Qy      5266 -CTTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAA 5324
Db     96667 TCCTTCCTTCCTTCCTTTCTCTTTCTTTTTTTGTCTTTCAGTGGAGTTGCAGCTCCTGAA 96608

Qy      5325 AGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTC 5384
Db     96607 A-CATCTGTGTTTATAAGCCTCATGTGTCATTCATGTATGACTCCACCTACATTTC---C 96552

Qy      5385 TCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATC 5444
Db     96551 TCCCCCCCACTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGTTTCGTATC 96492

Qy      5445 TCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCA 5504
Db     96491 TCGTTAAAGACAGGATTGTCACTTGAACTACTTCTGCAGCATTCAGAGTGGCCACGGCCA 96432

Qy      5505 ACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGT 5564
Db     96431 ACACCACCGGATGTTTATTCATTGCTCTGAAGGTCAAAAGTCTCATTTTGTTTTCCTGGT 96372

Qy      5565 TAGATT--------CTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCT 5616
Db     96371 TAGATTCCCCACCCCCAACCCCCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCT 96312

Qy      5617 CTTAGCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCT 5676
Db     96311 CTTAGCATCACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTC 96254

Qy      5677 GGGTTTAAATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAA 5736
Db     96253 TGGTTTAAATTAAACAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAA 96194

Qy      5737 TTG-TTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATA 5795
Db     96193 TTGTTTCTCAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATA 96134

Qy      5796 AGAAAGTAGCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGT 5855
Db     96133 AGAAAGTAGCCAGTACTCACTAGCGTTATATATGATTTGGGTTCTGAGTAACTGGGGAGT 96074

Qy      5856 GTTAGCTTTGTGACTTTGTAGCACCAGGTCTTATTAGGAA--AGTCTGTTGGCCTTTTAC 5913
Db     96073 GTTAGATTTGTGACTTTGTAGCAGCAGGTTTTATTAGGAAAGAGTCTATTGGCCTTTTAC 96014

Qy      5914 AGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTA 5973
Db     96013 AGGGTATTAATCCCTTTGTCGCTTACCATGGATGCCTTAAGTTCTTTTGAGTCTCATTTA 95954
```

```
Qy       5974 AGAA--TTCCTTTTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGTT 6031
              ||||  |  |||||||||||||||||| |||| | | | ||||||| ||||||||| ||
Db      95953 AGAACCTCTCTTTTCTCGAAGCATGACGAGTGCA-CTCGATACTTACTTGCTCACTCATT 95895

Qy       6032 TACCTG-GCTTAGTTTGTGCTGGGTTATTTAATTGCACTTTCCAGCATCATGCTTCCTCC 6090
              ||||||  | |  || |||||| ||||||  |  |||||||||||||||||||||||||
Db      95894 TACCTGTCCCTCCTTCGTGCTGGTTTATTTCCTCACACTTTCCAGCATCATGCTTCCTCC 95835

Qy       6091 TTACAAATATGATATTCTTTATTGTTACACTAAGGTGTTGATCATGTATCTGTCCCTGTA 6150
              ||||||||||||||||||||| |||||||||||||  ||||||||||||||||||| |||
Db      95834 TTACAAATATGATATTCTTTAGTGTTACACTAAGGCATTGATCATGTATCTGTCCCCGTA 95775

Qy       6151 AAGAATTAATAAACTATTTTCCAGAC 6176
              | ||||||||||||||||||||||||
Db      95774 ATGAATTAATAAACTATTTTCCAGAC 95749


RESULT 14
AC068507/c
LOCUS       AC068507               176282 bp    DNA     linear   HTG 06-MAY-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-349H15 map 15, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC068507
VERSION     AC068507.2  GI:8389522
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349H15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 9, 2000 this sequence version replaced gi:7684421.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10275
                Center clone name: 349_H_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156826 bases at least Q40
                Consensus quality: 166608 bases at least Q30
                Consensus quality: 170674 bases at least Q20
                Insert size: 173682; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      1477: contig of 1477 bp in length
            *       1478      1577: gap of 100 bp
            *       1578      2785: contig of 1208 bp in length
            *       2786      2885: gap of 100 bp
            *       2886      4859: contig of 1974 bp in length
            *       4860      4959: gap of 100 bp
            *       4960      6416: contig of 1457 bp in length
            *       6417      6516: gap of 100 bp
            *       6517      8739: contig of 2223 bp in length
            *       8740      8839: gap of 100 bp
            *       8840     10868: contig of 2029 bp in length
            *      10869     10968: gap of 100 bp
            *      10969     13737: contig of 2769 bp in length
            *      13738     13837: gap of 100 bp
            *      13838     15329: contig of 1492 bp in length
            *      15330     15429: gap of 100 bp
            *      15430     17938: contig of 2509 bp in length
            *      17939     18038: gap of 100 bp
            *      18039     19981: contig of 1943 bp in length
            *      19982     20081: gap of 100 bp
            *      20082     23587: contig of 3506 bp in length
            *      23588     23687: gap of 100 bp
            *      23688     27447: contig of 3760 bp in length
            *      27448     27547: gap of 100 bp
            *      27548     30161: contig of 2614 bp in length
            *      30162     30261: gap of 100 bp
            *      30262     34737: contig of 4476 bp in length
            *      34738     34837: gap of 100 bp
            *      34838     40101: contig of 5264 bp in length
            *      40102     40201: gap of 100 bp
            *      40202     44039: contig of 3838 bp in length
            *      44040     44139: gap of 100 bp
            *      44140     49950: contig of 5811 bp in length
            *      49951     50050: gap of 100 bp
            *      50051     55555: contig of 5505 bp in length
            *      55556     55655: gap of 100 bp
            *      55656     60496: contig of 4841 bp in length
            *      60497     60596: gap of 100 bp
            *      60597     68544: contig of 7948 bp in length
            *      68545     68644: gap of 100 bp
            *      68645     76940: contig of 8296 bp in length
            *      76941     77040: gap of 100 bp
            *      77041     85480: contig of 8440 bp in length
            *      85481     85580: gap of 100 bp
            *      85581     93905: contig of 8325 bp in length
            *      93906     94005: gap of 100 bp
            *      94006    108063: contig of 14058 bp in length
            *     108064    108163: gap of 100 bp
            *     108164    123496: contig of 15333 bp in length
            *     123497    123596: gap of 100 bp
            *     123597    139965: contig of 16369 bp in length
            *     139966    140065: gap of 100 bp
            *     140066    176282: contig of 36217 bp in length.
FEATURES             Location/Qualifiers
     source          1..176282
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349H15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1477
                     /note="assembly_fragment"
     misc_feature    1578..2785
                     /note="assembly_fragment"
     misc_feature    2886..4859
                     /note="assembly_fragment"
     misc_feature    4960..6416
                     /note="assembly_fragment"
     misc_feature    6517..8739
                     /note="assembly_fragment"
     misc_feature    8840..10868
                     /note="assembly_fragment"
     misc_feature    10969..13737
                     /note="assembly_fragment"
     misc_feature    13838..15329
                     /note="assembly_fragment"
     misc_feature    15430..17938
                     /note="assembly_fragment"
     misc_feature    18039..19981
                     /note="assembly_fragment"
     misc_feature    20082..23587
                     /note="assembly_fragment"
     misc_feature    23688..27447
                     /note="assembly_fragment"
     misc_feature    27548..30161
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    30262..34737
                     /note="assembly_fragment"
     misc_feature    34838..40101
                     /note="assembly_fragment"
     misc_feature    40202..44039
                     /note="assembly_fragment"
     misc_feature    44140..49950
                     /note="assembly_fragment"
     misc_feature    50051..55555
                     /note="assembly_fragment"
     misc_feature    55656..60496
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    60597..68544
                     /note="assembly_fragment"
     misc_feature    68645..76940
                     /note="assembly_fragment"
     misc_feature    77041..85480
                     /note="assembly_fragment"
     misc_feature    85581..93905
                     /note="assembly_fragment"
     misc_feature    94006..108063
                     /note="assembly_fragment"
     misc_feature    108164..123496
                     /note="assembly_fragment"
     misc_feature    123597..139965
                     /note="assembly_fragment"
     misc_feature    140066..176282
                     /note="assembly_fragment"
BASE COUNT    44272 a  43276 c  42765 g  43361 t   2608 others
ORIGIN

  Query Match             13.8%;  Score 854.8;  DB 2;  Length 176282;
  Best Local Similarity   64.5%;  Pred. No. 1.9e-193;
  Matches 1930;  Conservative    0;  Mismatches  847;  Indels  217;  Gaps   35;

Qy       3341 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 3400
Db     154550 AGGGCAATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCA 154491

Qy       3401 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 3460
Db     154490 TTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGG 154431

Qy       3461 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 3520
Db     154430 AGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAG 154371

Qy       3521 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 3580
Db     154370 GGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGC 154311

Qy       3581 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 3640
Db     154310 CAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCC 154251

Qy       3641 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 3700
Db     154250 AGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCC 154191

Qy       3701 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 3760
Db     154190 CTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCTGCCTAGCTCTTCC 154131

Qy       3761 CAGA----------------AGGCAGATATGGCTCAGGAA-------------CATGCCA 3791
Db     154130 CAGAGGATGTGGTTTGGGGCAGGCAGGTATGGATCACATAGGATGCGATACCTGTGGCCG 154071

Qy       3792 TGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGCCTCTTAGGGTC 3851
Db     154070 TGTATGTCCACATGTGTGCCTGTAGATACATCATCAAGCCCTTTGGAGCTTCCTAAGTTG 154011

Qy       3852 CTTTGGCTGGGGTTGGGGAGAA------CTTTACTCTCCCTCATATTCTGCATCACATAC 3905
Db     154010 CTTTGGCTGAGGGGAGAGGAAAACATGGATTATTCACTCCCCCCATACTCTTTGTGATAC 153951

Qy       3906 AGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGTGTGTG 3965
Db     153950 ACATGTGACATGTGAAAGACATACGAGACATAGCTACATGTGATGTGCACATGTGTGAAG 153891

Qy       3966 TGTGTGTGTG-----TGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
Db     153890 TGCATGTATGCGTACTGGTTGTTGAGCTGGGAAACCGTGGCCCAGGCAGTGGTCACTACA 153831

Qy       4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTG--TAACAGTGATCAGT-GCCAGCTCTTTGA 4077
Db     153830 GCCTGATTGGTCCTCCAGGTCAGAACGGTGCCCCACAGTGGTCAGTCCCCAGCCCTGTGG 153771

Qy       4078 GCTTTTAGCCTTGTCACCTAGCCTTTTATT--ACACTCTGAGAGTGTCTCCAGTGCTGTG 4135
Db     153770 GCCCCCACCTCCATCGCCCAGCCTTTTATTACACACTCTGAGAGTGTCTCCAATGCCTGT 153711

Qy       4136 TCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAAC 4195
Db     153710 CTGACAAAGACAGTCCCAGCCCATTCTC---CTGTCTGGCTGGGTTGGGTGCAA------ 153660

Qy       4196 TCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTA 4255
Db     153659 -------------------GCAGGCTCTGAATGCCTGGCATTTCAGCTGCATCACCTCCC 153619

Qy       4256 AGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACC 4315
Db     153618 AGCTCCTTATTGCCCAAATAGAGAGGGTGGCCCTGGCTCCCCTCCGAGCAACTCTGCATT 153559

Qy       4316 TAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCT 4375
Db     153558 TAATTTTGTAATCTGGGAAGTGCCTGGTTTTGAAAATCCGCTTTCTCTCACTCTTCCCCT 153499

Qy       4376 CTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCT-CATGTCATCTTGCTCGATTATC-C 4433
Db     153498 CCTTCCTTGCCCCTGGCTGCTCTAGTGTTCTGTCTCCCAGTCACCTCGCTCTCCCAGCAC 153439
```

```
Qy       4434 TGGGGCCCTTCTCTT--TCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATT---- 4487
Db     153438 CAGTGCCCTTCTCCTGCTCCCAGATACTCTTTCCTTTCCTCTCTCCTGTTTTCCTTCCTC 153379

Qy       4488 -----TCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCT 4542
Db     153378 TGCTATCTCTCACACCTCTCCCAGACTATGTCATCTTGTTCTCCTGCCTGGGTTCAAACT 153319

Qy       4543 ATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGA---------CCATAG 4593
Db     153318 CTGCATCCTTCTCTAACAACGTGACTACCTCATGTCTGCTTCAAGGCCCCCGTGCCCTTC 153259

Qy       4594 TGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTG--------------- 4638
Db     153258 CTGTATCCGCGGCTGCCGCGCACTCGCCTGCCATCCTCCTGCCTCCTCTTCACTCAGTGC 153199

Qy       4639 ------------------GGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTT 4679
Db     153198 TTCTGCTTGCCCTGCCCCAGGCAGCCCACCCACGCCCAGTGCGGGTGTGGAGAAGATCTT 153139

Qy       4680 GTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGA 4739
Db     153138 CTGGCTTCCCTGCAT--CTTGCCTTTGGGATTGGGATCCAAGGGTTCTCCATGGATG--- 153084

Qy       4740 TGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTG 4799
Db     153083 -GATCCAAGTCATAGAGGGGAATGTTTGAGACAGGGAAGGGGGCTGTGATCCAGAGGCTC 153025

Qy       4800 AAAAAAAATGGCTGCCACCC---ATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGC 4856
Db     153024 AGAATAAAAAGATGCCCTCCCTTCTATGCAGGGGGGCAAGTTTACTGGATGGAGATGATT 152965

Qy       4857 AGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAG 4916
Db     152964 TGGGCCTCTCTTCCAGAAGAAGCTAAAGGAAGAGAAGGGGAGTGAGAGTTCAGGGAGGCC 152905

Qy       4917 TTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTT------CCTCC 4970
Db     152904 CTTCCCACCCTGTGAGGCTTGACTTGATCTGGATTGGGGATGACAGGAATCTCACCCCTCT 152845

Qy       4971 AAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATC 5030
Db     152844 GGGGTGCTGGCAAGGAGGTCTTTGCACAGGAAAAGG--GGTAGCTCATTTCAGTTTGTTT 152787

Qy       5031 TTTCTTTACACTGAATCCCAAAATCAT----------CTTACCACAAAGGGCCAAGCCTG 5080
Db     152786 TTTCTTTAAATTGAATCCTCAAGTCATTTTCTGTTCACCTGCCGCACAGGGACAAGCTTG 152727

Qy       5081 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 5140
Db     152726 ACTTCTATTTTCTGTGTAGTGAAAACAATGTCATTTATTTGGTTTTTCACCTCAGCCCTC 152667

Qy       5141 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 5200
Db     152666 TCATAGGAGCATAGAATGTTAGGGTCTTTACTCCCT--------------AATGATGTC 152622

Qy       5201 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 5260
Db     152621 TGATTGGCACATCAAGAGTTAACTCTGCCTTCTGGGCCAAATTCGAAATAACCAGTCCAT 152562

Qy       5261 TTTCCC--TTTATTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCT 5318
Db     152561 TTTTCCTTTTTTTTTTTTTTTTTTTTTAAATGGTGGAATGTCTCTCAGCACAGTTGCGGCT 152502

Qy       5319 T---------CTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGA---C 5366
Db     152501 TCCTCAAACCCTGAAAGCATCTGTGTTTATTATACTCGGGTGTCACTCACTGTTGATGTC 152442

Qy       5367 CCCACCCACATTTCCTTCTCCTCCCC--TCTTCAGCCAGCCTATGATAACACTAAAGATT 5424
Db     152441 TGCACCTACGTTTCCACCTCCTCCCCCTCCTTCAGCCAGCCTATGATAACACTAAAGATT 152382

Qy       5425 ATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGC 5484
Db     152381 ATTAATGTTGGTTTTGTATCTCGTTAAAGACAGAATTGTCACTTGTAGTATTTCTGTAGC 152322

Qy       5485 ATTCAAAGTGGCCACGGCCAACACCACC--GTATGTTTCTTCATTGCTCTGAAGGTCAAG 5542
Db     152321 ATTCAGCGCTGCTGTGGCTAACACCACTGTGTATGTTTCATCATTGCTCTGAAGGTCAAA 152262

Qy       5543 AGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCT----------------CCTTGCCT 5586
Db     152261 AGCCTCATTTTATTTTGCTGGTTTGATTTTTTTTTTTAAAGAAGAAAAAAAAACTGCCC 152202

Qy       5587 TGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCA 5646
Db     152201 TGAATTAAATGGCTGTTTTAACAGTAGGCTCTTAGCATTATACCACATAGTCATTTTTCA 152142

Qy       5647 TGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGAC 5706
Db     152141 TGTTCTTGTTTAACAGGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGAC 152084

Qy       5707 AATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAAT 5765
Db     152083 AATTTATAACCCATTAGGTTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAAT 152024

Qy       5766 CTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATA 5825
Db     152023 CTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTGGCCATTTCTCACTAGCACTATA 151964

Qy       5826 TATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTC 5885
Db     151963 TATGATTTGGGCTCTGGGTAATTTGGAAGTGTTAGGTTTGTGTCTTTGTAGCAGTATTTT 151904

Qy       5886 T-TATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGG 5944
Db     151903 TATTAGAAAAGAATCTATTGGCCTTTTACAGGGTATTAATCCCTTTGTCACCTACCATTG 151844

Qy       5945 ATGCCTTAAGTTCTTTG-GAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAAGTG 6003
Db     151843 ATGCCTTAAGTTTTCTGAGTCTCAATTAAAAATCTTCCTTTTCTTGATGCATGACAAGTG 151784

Qy       6004 TATCGCAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGTGCTGGGTTATTTAA 6062
Db     151783 T-----AATCAGTACTTGCTCATTTATTTGTCTGTATTTAGTTTATGCTGTACTATTTAA 151729

Qy       6063 TTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACACTA 6122
Db     151728 TTATCCTTCCAGCGTTTTTTTTTCTCCTTACAAATATGATACTCTTTAGTGTTAAGCTA 151669

Qy       6123 AGGTGTTGA-TCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGA 6175
Db     151668 AGGCATTGATTCATGTATCTGTCCTTATAATGAATTAATAAACTATTTTCCAGA 151615
```

RESULT 15
AC105129

```
LOCUS       AC105129               198295 bp    DNA     linear   PRI 26-APR-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-349G13, complete sequence.
ACCESSION   AC105129
VERSION     AC105129.4  GI:20330981
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
```

```
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 26, 2002 this sequence version replaced gi:20148056.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22971
                Center clone name: 349_G_13
            --------------.
FEATURES             Location/Qualifiers
     source          1..198295
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1..79)
                     /rpt_family="AluS"
     repeat_region   complement(183..494)
                     /rpt_family="AluSq"
     repeat_region   complement(1216..1416)
                     /rpt_family="L1MB8"
     repeat_region   complement(1417..1709)
                     /rpt_family="AluSg"
     repeat_region   complement(1710..2115)
                     /rpt_family="L1MB8"
     repeat_region   2116..2169
                     /rpt_family="(TTTTA)n"
     repeat_region   2186..2234
                     /rpt_family="AluSq/x"
     repeat_region   complement(2236..2512)
                     /rpt_family="AluJb"
     repeat_region   complement(2514..2631)
                     /rpt_family="L1MB8"
     repeat_region   complement(2632..2941)
                     /rpt_family="AluY"
     repeat_region   complement(2942..3104)
                     /rpt_family="L1MB8"
     repeat_region   3181..3464
                     /rpt_family="AluSx"
     repeat_region   3465..3711
                     /rpt_family="AluJb"
     repeat_region   3712..3788
                     /rpt_family="(TA)n"
     repeat_region   complement(4661..4940)
                     /rpt_family="AluJb"
     repeat_region   complement(4946..5106)
                     /rpt_family="FRAM"
     repeat_region   complement(5878..6193)
                     /rpt_family="AluYb8"
     repeat_region   7197..7490
                     /rpt_family="AluSx"
     repeat_region   7925..8226
                     /rpt_family="AluY"
     repeat_region   complement(8227..8312)
                     /rpt_family="L1MB3"
     repeat_region   8313..8347
                     /rpt_family="AT_rich"
     repeat_region   complement(8350..8617)
                     /rpt_family="AluSx"
     repeat_region   complement(8621..9274)
                     /rpt_family="L1MB3"
     repeat_region   9403..9445
                     /rpt_family="AT_rich"
     repeat_region   9706..10007
                     /rpt_family="AluJb"
     repeat_region   complement(10940..11000)
                     /rpt_family="L1MB5"
```

```
repeat_region   11001. .11284
                /rpt_family="AluJo"
repeat_region   complement(11285. .11366)
                /rpt_family="L1MB5"
repeat_region   complement(11641. .11934)
                /rpt_family="AluSg"
repeat_region   complement(12071. .12294)
                /rpt_family="AluJb"
repeat_region   complement(12410. .12697)
                /rpt_family="AluSx"
repeat_region   complement(12708. .12998)
                /rpt_family="AluJb"
repeat_region   13158. .13478
                /rpt_family="AluJo"
repeat_region   13970. .14093
                /rpt_family="FLAM_C"
repeat_region   14108. .14412
                /rpt_family="AluJo"
repeat_region   complement(14813. .15119)
                /rpt_family="AluSx"
repeat_region   complement(15130. .15432)
                /rpt_family="AluSq"
repeat_region   complement(15532. .15841)
                /rpt_family="AluSx"
repeat_region   complement(15906. .16077)
                /rpt_family="AluSg/x"
repeat_region   complement(16079. .16383)
                /rpt_family="AluSq"
repeat_region   16409. .16481
                /rpt_family="MER5A"
repeat_region   16482. .16779
                /rpt_family="AluSq"
repeat_region   16780. .16874
                /rpt_family="MER5A"
repeat_region   complement(17247. .17543)

Query Match            13.8%;  Score 854.8;  DB 9;  Length 198295;
Best Local Similarity  64.5%;  Pred. No. 1.9e-193;
Matches 1930;  Conservative    0;  Mismatches  847;  Indels  217;  Gaps   35;

Qy       3341 AGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCA 3400
Db     101335 AGGGCAATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCA 101394

Qy       3401 TTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGG 3460
Db     101395 TTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGG 101454

Qy       3461 AGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGT 3520
Db     101455 AGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAG 101514

Qy       3521 CTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGC 3580
Db     101515 GGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGC 101574

Qy       3581 CAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCC 3640
Db     101575 CAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCC 101634

Qy       3641 AGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTC 3700
Db     101635 AGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCC 101694

Qy       3701 CTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAG 3760
Db     101695 CTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCTGCCTAGCTCTTCC 101754

Qy       3761 CAGA----------------AGGCAGATATGGCTCAGGAA-------------CATGCCA 3791
Db     101755 CAGAGGATGTGGTTTGGGGCAGGCAGGTATGGATCACATAGGATGCGATACCTGTGGCCG 101814

Qy       3792 TGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGCCTCTTAGGGTC 3851
Db     101815 TGTATGTCCACATGTGTGCCTGTAGATACATCATCAAGCCCTTTGGAGCTTCCTAAGTTG 101874

Qy       3852 CTTTGGCTGGGGTTGGGGAGAA------CTTTACTCTCCCTCATATTCTGCATCACATAC 3905
Db     101875 CTTTGGCTGAGGGGAGAGGAAAACATGGATTATTCACTCCCCCATACTCTTTGTGATAC 101934

Qy       3906 AGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGTGTGTG 3965
Db     101935 ACATGTGACATGTGAAAGACATACGAGACATAGCTACATGTGATGTGCACATGTGTGAAG 101994

Qy       3966 TGTGTGTGTG-----TGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
Db     101995 TGCATGTATGCGTACTGGTTGTTGAGCTGGGAAACCGTGGCCCAGGCAGTGGTCACTACA 102054

Qy       4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTG--TAACAGTGATCAGT-GCCAGCTCTTTGA 4077
Db     102055 GCCTGATTGGTCCTCCAGGTCAGAACGGTGCCCCACAGTGGTCAGTCCCCAGCCCTGTGG 102114

Qy       4078 GCTTTTAGCCTTGTCACCTAGCCTTTTATT--ACACTCTGAGAGTGTCTCCAGTGCTGTG 4135
Db     102115 GCCCCCACCTCCATCGCCCAGCCTTTTATTACACACTCTGAGAGTGTCTCCAATGCCTGT 102174

Qy       4136 TCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAAC 4195
Db     102175 CTGACAAAGACAGTCCCAGCCCATTCTC---CTGTCTGGCTGGGTTGGGTGCAA------ 102225

Qy       4196 TCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTA 4255
Db     102226 ------------------GCAGGCTCTGAATGCCTGGCATTTCAGCTGCATCACCTCCC 102266

Qy       4256 AGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACC 4315
Db     102267 AGCTCCTTATTGCCCAAATAGAGAGGGTGGCCCTGGCTCCCCTCCGAGCAACTCTGCATT 102326

Qy       4316 TAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCT 4375
Db     102327 TAATTTTGTAATCTGGGAAGTGCCTGGTTTTGAAAATCCGCTTTCTCTCACTCTTCCCCT 102386

Qy       4376 CTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCT-CATGTCATCTTGCTCGATTATC-C 4433
Db     102387 CCTTCCTTGCCCCTGGCTGCTCTAGTGTTCTGTCTCCCAGTCACCTCGCTCTCCCAGCAC 102446

Qy       4434 TGGGGCCCTTCTCTT--TCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATT---- 4487
Db     102447 CAGTGCCCTTCTCCTGCTCCCAGATACTCTTTCCTTTCCTCTCTCCTGTTTTCCTTCCTC 102506

Qy       4488 -----TCTGTCTGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCT 4542
Db     102507 TGCTATCTCTCACACCTCTCCCAGACTATGTCATCTTGTTCTCCTGCCTGGGTTCAAACT 102566

Qy       4543 ATGCACCCTTTCCTAACAACATGACTACCTCATGTCTGCTTCAGA---------CCATAG 4593
Db     102567 CTGCATCCTTCTCTAACAACGTGACTACCTCATGTCTGCTTCAAGGCCCCCGTGCCCTTC 102626

Qy       4594 TGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACCGCCTTCCTG--------------- 4638
Db     102627 CTGTATCCGCGGCTGCCGCGCACTCGCCTGCCATCCTCCTGCCTCCTCTTCACTCAGTGC 102686

Qy       4639 ------------------GGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTT 4679
Db     102687 TTCTGCTTGCCCTGCCCCAGGCAGCCCACCCACGCCCAGTGCGGGTGTGGAGAAGATCTT 102746

Qy       4680 GTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGA 4739
Db     102747 CTGGCTTCCCTGCAT--CTTGCCTTTGGGATTGGGATCCAAGGGTTCTCCATGGATG--- 102801

Qy       4740 TGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTG 4799
Db     102802 -GATCCAAGTCATAGAGGGGAATGTTTGAGACAGGGAAGGGGGCTGTGATCCAGAGGCTC 102860

Qy       4800 AAAAAAAATGGCTGCCACCC---ATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGC 4856
```

```
Db    102861 AGAATAAAAAGATGCCCTCCCTTCTATGCAGGGGGGCAAGTTTACTGGATGGAGATGATT 102920

Qy      4857 AGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAG 4916
Db    102921 TGGGCCTCTCTTCCAGAAGAAGCTAAAGGAAGAGAAGGGGAGTGAGAGTTCAGGGAGGCC 102980

Qy      4917 TTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTT------CCTCC 4970
Db    102981 CTTCCCACCCTGTGAGGCTTGACTTGATCTGGATTGGGGATGACAGGAATCTCACCCTCT 103040

Qy      4971 AAGATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATC 5030
Db    103041 GGGGTGCTGGCAAGGAGGTCTTTGCACAGGAAAAGG--GGTAGCTCATTTCAGTTTGTTT 103098

Qy      5031 TTTCTTTACACTGAATCCCAAAATCAT----------CTTACCACAAAGGGCCAAGCCTG 5080
Db    103099 TTTCTTTAAATTGAATCCTCAAGTCATTTTCTGTTCACCTGCCGCACAGGGACAAGCTTG 103158

Qy      5081 ACTGGTATTTCCTGAGTCACAAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATG 5140
Db    103159 ACTTCTATTTTCTGTGTAGTGAAAACAATGTCATTTATTTGGTTTTTCACCTCAGCCCTC 103218

Qy      5141 TCCCAGAATTGTCAGATCCAGTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAAC 5200
Db    103219 TCATAGGAGCATAGAATGTTAGGGTCTTTACTCCCT---------------AATGATGTC 103263

Qy      5201 TGGCTGGCACATCAAGTGTTAACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTAT 5260
Db    103264 TGATTGGCACATCAAGAGTTAACTCTGCCTTCTGGGCCAAATTCGAAATAACCAGTCCAT 103323

Qy      5261 TTTCCC--TTTATTTTATTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCT 5318
Db    103324 TTTTCCTTTTTTTTTTTTTTTTTTTTAAATGGTGGAATGTCTCTCAGCACAGTTGCGGCT 103383

Qy      5319 T---------CTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGA---C 5366
Db    103384 TCCTCAAACCCTGAAAGCATCTGTGTTTATTATACTCGGGTGTCACTCACTGTTGATGTC 103443

Qy      5367 CCCACCCACATTTCCTTCTCCTCCCC--TCTTCAGCCAGCCTATGATAACACTAAAGATT 5424
Db    103444 TGCACCTACGTTTCCACCTCCTCCCCCTCCTTCAGCCAGCCTATGATAACACTAAAGATT 103503

Qy      5425 ATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGC 5484
Db    103504 ATTAATGTTGGTTTTGTATCTCGTTAAAGACAGAATTGTCACTTGTAGTATTTCTGTAGC 103563

Qy      5485 ATTCAAAGTGGCCACGGCCAACACCACC--GTATGTTTCTTCATTGCTCTGAAGGTCAAG 5542
Db    103564 ATTCAGCGCTGCTGTGGCTAACACCACTGTGTATGTTTCATCATTGCTCTGAAGGTCAAA 103623

Qy      5543 AGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCT----------------CCTTGCCT 5586
Db    103624 AGCCTCATTTTATTTTGCTGGTTTGATTTTTTTTTTTAAAGAAGAAAAAAAAACTGCCC 103683

Qy      5587 TGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCA 5646
Db    103684 TGAATTAAATGGCTGTTTTAACAGTAGGCTCTTAGCATTATACCACATAGTCATTTTTCA 103743

Qy      5647 TGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGAC 5706
Db    103744 TGTTCTTGTTTAACAGGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGAC 103801

Qy      5707 AATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAAT 5765
Db    103802 AATTTATAACCCATTAGGTTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAAT 103861

Qy      5766 CTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATA 5825
Db    103862 CTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTGGCCATTTCTCACTAGCACTATA 103921

Qy      5826 TATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTC 5885
Db    103922 TATGATTTGGGCTCTGGGTAATTTGGAAGTGTTAGGTTTGTGTCTTTGTAGCAGTATTTT 103981

Qy      5886 T-TATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGG 5944
Db    103982 TATTAGAAAAGAATCTATTGGCCTTTTACAGGGTATTAATCCCTTTGTCACCTACCATTG 104041

Qy      5945 ATGCCTTAAGTTCTTTG-GAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAAGTG 6003
Db    104042 ATGCCTTAAGTTTTCTGAGTCTCAATTAAAAATCTTCCTTTTCTTGATGCATGACAAGTG 104101

Qy      6004 TATCGCAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGTGCTGGGTTATTTAA 6062
Db    104102 T-----AATCAGTACTTGCTCATTTATTTGTCTGTATTTAGTTTATGCTGTACTATTTAA 104156

Qy      6063 TTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACACTA 6122
Db    104157 TTATCCTTCCAGCGTTTTTTTTTCTCCTTACAAATATGATACTCTTTAGTGTTAAGCTA 104216

Qy      6123 AGGTGTTGA-TCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGA 6175
Db    104217 AGGCATTGATTCATGTATCTGTCCTTATAATGAATTAATAAACTATTTTCCAGA 104270
```

Search completed: October 15, 2003, 04:35:49
Job time : 23145 secs

```
OM nucleic - nucleic search, using sw model

Run on:         October 14, 2003, 17:04:36 ; Search time 14061.8 Seconds
                                            (without alignments)
                                            10674.595 Million cell updates/sec

Title:          US-09-754-997A-1
Perfect score:  6176
Sequence:       1 atggcgcgggcggacacggg..........taataaactattttccagac 6176

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  em_gss_hum:*
                 18:  em_gss_inv:*
                 19:  em_gss_pln:*
                 20:  em_gss_vrt:*
                 21:  em_gss_fun:*
                 22:  em_gss_mam:*
                 23:  em_gss_mus:*
                 24:  em_gss_pro:*
                 25:  em_gss_rod:*
                 26:  em_gss_phg:*
                 27:  em_gss_vrl:*
                 28:  gb_gss1:*
                 29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1479	23.9	2626	11	AK079073	AK079073 Mus muscu
	2	883	14.3	947	11	AK020142	AK020142 Mus muscu
	3	805.8	13.0	849	14	CD352600	CD352600 UI-M-GI0-
	4	792.2	12.8	817	14	CA317106	CA317106 UI-M-FW0-
	5	737.6	11.9	748	14	CD348071	CD348071 UI-M-FY0-
	6	726.6	11.8	741	14	CD352400	CD352400 UI-M-GI0-
	7	720.6	11.7	944	13	BU840589	BU840589 AGENCOURT
	8	714.8	11.6	808	13	BU705518	BU705518 UI-M-FO0-
	9	706.4	11.4	719	14	CA751384	CA751384 UI-M-FO0-
	10	703.8	11.4	740	13	BU701299	BU701299 UI-M-EX0-
	11	690	11.2	713	14	CB524810	CB524810 UI-M-FY0-
	12	685.8	11.1	699	14	CD351162	CD351162 UI-M-FY0-
	13	639.6	10.4	900	12	BI185886	BI185886 UNL-P-FN-
	14	629.4	10.2	631	14	CB518504	CB518504 UI-M-GH0-
	15	618.2	10.0	671	14	BY719391	BY719391 BY719391
	16	617.2	10.0	624	13	BU056331	BU056331 UI-M-FO0-
	17	599.8	9.7	651	10	BB624536	BB624536 BB624536
	18	582	9.4	591	13	BU057120	BU057120 UI-M-FO0-
	19	581.6	9.4	637	10	BB665894	BB665894 BB665894
	20	580	9.4	630	10	BB626002	BB626002 BB626002
c	21	574	9.3	621	9	AW049847	AW049847 UI-M-BH1-
	22	573	9.3	660	10	BB071741	BB071741 BB071741
	23	561	9.1	583	14	CA895918	CA895918 B0195D06-
	24	560.4	9.1	632	10	BB649997	BB649997 BB649997
	25	556.4	9.0	759	14	CB988160	CB988160 AGENCOURT
	26	555.4	9.0	579	28	AZ746941	AZ746941 RPCI-24-1
	27	548.8	8.9	553	14	CA316452	CA316452 UI-M-FW0-
	28	546.2	8.8	703	14	BY712424	BY712424 BY712424
	29	541.6	8.8	584	28	AZ745366	AZ745366 RPCI-24-1
	30	538.4	8.7	1137	11	AK082162	AK082162 Mus muscu
	31	534.4	8.7	544	14	CA892088	CA892088 B0170A06-
	32	529	8.6	551	13	BU705885	BU705885 UI-M-FR0-
	33	513.2	8.3	578	14	CB607425	CB607425 AMGNNUC:N
	34	509	8.2	509	14	CA873325	CA873325 K0924D02-
c	35	508.8	8.2	723	14	CA510263	CA510263 UI-R-FS0-
	36	505	8.2	528	4	BX519858	Bx519858 RZPD Mus
c	37	503.6	8.2	717	13	BX100607	BX100607 BX100607
c	38	502	8.1	565	14	CA883712	CA883712 B0105A08-
	39	501.6	8.1	542	4	BX523249	Bx523249 RZPD Mus
	40	500.8	8.1	504	14	CA876082	CA876082 K0945H09-
c	41	491.6	8.0	553	14	CA871246	CA871246 K0908H04-
	42	485.6	7.9	528	13	BU580111	BU580111 ik75c07.y
	43	462.4	7.5	809	29	BZ222348	BZ222348 CH230-372
	44	453.4	7.3	455	14	CA883713	CA883713 B0105A08-
	45	452	7.3	533	10	BB681084	BB681084 BB681084

ALIGNMENTS

```
RESULT 1
AK079073
LOCUS       AK079073                2626 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male diencephalon cDNA, RIKEN full-length
            enriched library, clone:9330155G14 product:hypothetical protein,
            full insert sequence.
ACCESSION   AK079073
VERSION     AK079073.1  GI:26098261
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
```

```
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2626)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2626
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9330155G14"
                     /db_xref="taxon:10090"
                     /clone="9330155G14"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1..2626
                     /note="hypothetical protein (evidence: rsCDS)"
BASE COUNT      599 a    767 c    724 g    536 t
ORIGIN

  Query Match             23.9%;  Score 1479;  DB 11;  Length 2626;
  Best Local Similarity   89.9%;  Pred. No. 3.4e-238;
  Matches 1734;  Conservative    0;  Mismatches  150;  Indels   45;  Gaps   12;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         15 ATGGCGCGGGCGGACACGGGCCGAGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 74

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 134

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 194

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 254

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 314

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 374

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 434

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 494

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 554

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
              ||||||||||||||| ||||||||||||||||||||||  |  |||||||||||||||||
Db        555 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 614

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              |||||||||||||||||||||||||||||||||||| ||||||||||||||| |||||||
Db        615 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGCCCACCAATTCAGCCCCCCAACGA 674

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              |||||||||||||||||||||||||||||||| | | ||||||||| || || || ||| 
Db        675 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTTAAAAGGTCTTTGGGGGGTAACAAGGGA 734

Qy        721 CAGGATGTGGTCATTGTG-GCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGT 779
              ||||||||||||||||||  |||||||||||||||||||||||||||||||||||| |||
Db        735 CAGGATGTGGTCATTGTGCCCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGT 794

Qy        780 AGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGA 839
```

```
Db        795 AGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGA 854

Qy        840 TGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAG 899

Db        855 TGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCCCCAG 914

Qy        900 CGCGCAGCCTCGGCACTCTGGAGTCT-ATGTCTGCCGAGCCAACAAGCCCCTCACGCGTG 958

Db        915 CGCCCAGCCTCGGCACTTTTGAATTTAATTTTTCCCAACCCACCAACCCCCCCACGCCTT 974

Qy        959 ACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGG--C 1016

Db        975 AATTTCCCCCTTCCGCTTCTTAAGTTCCAATTCTTTCTTCCCCCACCCTTTTCCCAGCCC 1034

Qy       1017 ACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGG 1076

Db       1035 GCCCAAGCCCTTTTCCCGAACCCGGCCCACCCCCCCCCGTTTCTTTTCCCGGCCTTCCGG 1094

Qy       1077 GGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCG 1136

Db       1095 GGACCCCCGGCCCGCGCTGCACTGGCTGCACGACGGAATCCCGTTGCGACCCAATGGGCG 1154

Qy       1137 CGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGC 1196

Db       1155 CGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGC 1214

Qy       1197 TGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCT 1256

Db       1215 TGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCT 1274

Qy       1257 GGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCT 1316

Db       1275 GGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCT 1334

Qy       1317 GAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAG-TTGCACAGCGAGCAAATCA 1375

Db       1335 GAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTTGCACAGCGAGCAAATCA 1394

Qy       1376 TTGGCTT--CTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATG--TGGAGTACCAG 1431

Db       1395 TTGGCTTTCTCTCTTTCACTACCAAAAGGCAAGGGGAGTGGACAATGGTGGGAGTACCAG 1454

Qy       1432 TTTGCAGTAAACAATGACACCACAGA----------------------GCTGCAGGTTCG 1469

Db       1455 TTTGCAGTAAAACAAAATGGGAAAACCAAAAGAAGCTTGGAAGGGTTTCGGGGAACCTGG 1514

Qy       1470 GGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGC 1529

Db       1515 GAAAACCCAACCCCGGATTAATGAAGTTTTTAAGTGGTTGGCCTAACTTCCAAGCTGGGG 1574

Qy       1530 CAGCC-------GAACCTCCAGCCCAG--CCCTGGTGCATACACTGGACGATGTCCCCAG 1580

Db       1575 GGGCCAGGCCGAAACCTTCAAGCCCAGGCCCCTGGTGCATACACTGGACGATGTCCCCAG 1634

Qy       1581 CGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATC-AGGGTGGCATGGC 1639

Db       1635 CGCACCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCCAGGGTGGCATGGC 1694

Qy       1640 TGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTT 1699

Db       1695 TGCCCCTGCCCTCCAGCATGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTT 1754

Qy       1700 TGGGGAAG---GAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTA 1756

Db       1755 TGGGGAAGGAAGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAAACACAACTTA 1814

Qy       1757 CGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCG 1816

Db       1815 CGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCG 1874

Qy       1817 CT-GGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAG 1875

Db       1875 CTGGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGGACAACCAG 1934

Qy       1876 AGCCATGTT 1884
              |||||||||
Db       1935 AGCCATGTT 1943
```

```
RESULT 2
AK020142
LOCUS       AK020142                 947 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 12 days embryo male wolffian duct includes surrounding
            region cDNA, RIKEN full-length enriched library, clone:6720465A10
            product:hypothetical protein, full insert sequence.
ACCESSION   AK020142
VERSION     AK020142.1  GI:12860634
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
```

```
            Group Phase I & II Team.
 TITLE      Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
 JOURNAL    Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 947)
 AUTHORS    Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 TITLE      Direct Submission
 JOURNAL    Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6720465A10"
                     /db_xref="MGI:1911600"
                     /db_xref="taxon:10090"
                     /clone="6720465A10"
                     /sex="male"
                     /tissue_type="wolffian duct includes surrounding region"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     misc_feature    1..947
                     /note="hypothetical protein (evidence: rsCDS)"
                     /db_xref="MGI:1858497"
BASE COUNT      241 a     190 c     168 g     348 t
ORIGIN

  Query Match             14.3%;  Score 883;  DB 11;  Length 947;
  Best Local Similarity   98.4%;  Pred. No. 3.9e-138;
  Matches  934;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    4;

Qy       5230 TTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTATTTATTTATTTATTT 5289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTATTTATTTATTTATTT 60

Qy       5290 ATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTG 5349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTG 120

Qy       5350 TGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATG 5409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATG 180

Qy       5410 ATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTG 5469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTG 240

Qy       5470 AACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGC 5529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGC 300

Qy       5530 TCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCCTTGCCTTGA 5589
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        301 TCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTTTTTTCCTCCTTGCCTTGA 360

Qy       5590 ATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCATGT 5649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCATGT 420

Qy       5650 TCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGACAAT 5709
              ||||||||||||||||||    | |   ||||||||||||||||||||||||||||||||
Db        421 TCTTGTTTAACAAGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGACAAT 478

Qy       5710 TTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAATCTG 5768
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        479 TTATAACCCATTAGGCTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAATCTG 538

Qy       5769 GTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATATAT 5828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 GTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATATAT 598

Qy       5829 GATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTCTTA 5888
              |||| |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        599 GATTTGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCAGCAGGTCTTA 658

Qy       5889 TTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGC 5948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 TTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGATGC 718

Qy       5949 CTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAAGTGTATCG 6008
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| |
Db        719 CTTAAGTTCTTTGGAGTCTCATTTAAGAATTTCTTTTCTCGAAGCATGACAAGTGTAT-G 777

Qy       6009 CAATACTTACATGCTCACTCGTTTACCTG-GCTTAGTTTGTGCTGGGTTATTTAATTGCA 6067
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        778 CAATACTTACATGCTCACTCGTTTACCTGTGCTTAGTTTGTGCTGGGTTATTTAATTGCA 837

Qy       6068 CTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACACTAAGGTG 6127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 CTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACACTAAGGTG 897

Qy       6128 TTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 TTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 946


RESULT 3
CD352600
LOCUS       CD352600                 849 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-GI0-cgk-h-06-0-UI.r1 NIH_BMAP_GI0 Mus musculus cDNA clone
            IMAGE: 6856471 5', mRNA sequence.
ACCESSION   CD352600
VERSION     CD352600.1  GI:31144111
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..849
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6856471"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GI0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector . The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      176 a    225 c    204 g    237 t      7 others
ORIGIN

  Query Match             13.0%;  Score 805.8;  DB 14;  Length 849;
  Best Local Similarity   98.6%;  Pred. No. 3.5e-125;
  Matches  840;  Conservative    0;  Mismatches    9;  Indels    3;  Gaps    3;

Qy       4390 ATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCCCTTCTCTTT 4449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCCCTTCTCTTT 60

Qy       4450 CCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATGCTTGTCTTT 4509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATGCTTGTCTTT 120

Qy       4510 ATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACAACATGACTA 4569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACAACATGACTA 180

Qy       4570 CCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACC 4629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCCCTGCCAACC 240

Qy       4630 GCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTGTGGCTTGTC 4689
              |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||
Db        241 GCCTTCCTGGGCAGATGA-CCCACTCCAAGTAGATCTGGAAAAGA-CCTTGTGGCTTGTC 298

Qy       4690 TGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGG 4749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGATGAGTCCAGG 358

Qy       4750 CTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATG 4809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 CTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGAAAAAAAATG 418

Qy       4810 GCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAA 4869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTCCTGGCTGAA 478

Qy       4870 ACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTG 4929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCCCTGACCCTG 538

Qy       4930 GAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGTAAGGGAAGT 4989
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        539 GAGCCCTGAACCTGGCTTCAGAGTAGCANAGAGTTTCCTCCAAGATGCTGTAAGGGAAGT 598

Qy       4990 CTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACACTGAATCCC 5049
              ||||||||||||||||||||||||||||||| |||| |||||||||||||||||||||||
Db        599 CTTTGCATAGGAAAAGGGCGGCTGGCTCATTNTATTNTATCTTTCTTTACACTGAATCCC 658

Qy       5050 AAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCAT 5109
              | ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        659 ANAATCATCTTACCACANAGGGCCAAGCCTGACTGGTATTTCCTGAGTCACAAGAGCCAT 718

Qy       5110 GCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCT 5169
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCATCTCTCTGGNTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCAGTGGCATCT 778

Qy       5170 GTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTTAACTCTGGC 5229
              |||||||||||||||||||||||||||||||||||| |||||||||||  | ||||||||
Db        779 GTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCT-GCACATCAAGTNGTTACTCTGGC 837

Qy       5230 TTCTGGGCCAAG 5241
              ||||||||||||
Db        838 TTCTGGGCCAAG 849


RESULT 4
CA317106
LOCUS       CA317106               817 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FW0-cbm-d-21-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
            IMAGE: 6811462 5', mRNA sequence.
ACCESSION   CA317106
VERSION     CA317106.1  GI:24535230
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..817
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6811462"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FW0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
```

```
              gel. First strand cDNA synthesis was primed with oligo-dT
              primer containing a Not I site. Double strand cDNA was
              size selected according to mRNA size fraction, ligated
              with EcoR I adaptor, digested with NotI and then cloned
              directionally into pYX-Asc vector. The library tag
              sequence located between the Not I site and the polyA tail
              is AGCGAGACAG. This library was created for the University
              Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
              Developing Mouse Nervous System', supported by National
              Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
              program coordinator."
BASE COUNT     180 a    254 c    234 g    148 t      1 others
ORIGIN

  Query Match             12.8%;  Score 792.2;  DB 14;  Length 817;
  Best Local Similarity   98.8%;  Pred. No. 6.8e-123;
  Matches  808;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy       3009 AGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCC 3068
              |||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCC 60

Qy       3069 TCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAG 3128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAG 120

Qy       3129 CCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGC 3188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGC 180

Qy       3189 TCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAG 3248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAG 240

Qy       3249 TGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACT 3308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACT 300

Qy       3309 GCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCC 3368
              |||||||||||||||||| ||||||| |||||||||||||||||||||||||||||||||
Db        301 GCTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCC 360

Qy       3369 AGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAA 3428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAA 420

Qy       3429 AGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGAC 3488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGAC 480

Qy       3489 TCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGA 3548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGA 540

Qy       3549 GCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGC 3608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGC 600

Qy       3609 AGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCC 3668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCC 660

Qy       3669 TGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAG 3728
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        661 TGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAG-CTTCCCAG 719

Qy       3729 GGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCAGGAACATG 3788
               |||||||||||||||||||| |  ||||||||||||||||||||||||||| |||||||
Db        720 NGCCCCTGTCTCCTCTGCTCAAGGTCCCTGAGCAGAAGGCAGATATGGCTCAAGAACATG 779

Qy       3789 CCATGCATGGCTACACATGTGTGTACTAGAGATATCCA 3826
              ||||||||||||||||||||||||||||||||||||||
Db        780 CCATGCATGGCTACACATGTGTGTACTAGAGATATCCA 817
```

```
RESULT 5
CD348071
LOCUS       CD348071               748 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-FY0-cfm-h-04-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6851477 5', mRNA sequence.
ACCESSION   CD348071
VERSION     CD348071.1  GI:31139586
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
             The following repetitive elements were found in this cDNA
            sequence: 179-203, >(CA)n#Simple_repeat (matched compliment)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..748
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6851477"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT     129 a    202 c    177 g    236 t      4 others
ORIGIN

  Query Match             11.9%;  Score 737.6;  DB 14;  Length 748;
  Best Local Similarity   98.9%;  Pred. No. 1e-113;
  Matches  740;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3775 GGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCT 3834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCTCAGGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCT 60

Qy       3835 TGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCT 3894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGAGCCTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCT 120

Qy       3895 GCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGT 3954
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGT 180

Qy       3955 GTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTC 4014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGTGTGTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTC 240

Qy       4015 ACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTT 4074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTT 300

Qy       4075 TGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGT 4134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGAGCTTTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGT 360

Qy       4135 GTCTACAAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAA 4194
              |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db        361 GTCTACAAAGACAGCGCCAGCCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAA 420

Qy       4195 CTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCT 4254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCACGGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCT 480

Qy       4255 AAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTAC 4314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTAC 540

Qy       4315 CTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTC 4374
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        541 CTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCNCTC 600

Qy       4375 TCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCT 4434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCT 660

Qy       4435 GGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCT 4494
              | |||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||
Db        661 GNGGCCCTTCTCTTTCNCATGATNGCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCT 720

Qy       4495 GCCATGCTTGTCTTTATGTCGTGTGTTT 4522
              ||||||||||||||||||||||||||||
Db        721 GCCATGCTTGTCTTTATGTCGTGTGTTT 748
```

```
RESULT 6
CD352400
LOCUS       CD352400                 741 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-GI0-cgj-p-07-0-UI.r1 NIH_BMAP_GI0 Mus musculus cDNA clone
            IMAGE: 6856280 5', mRNA sequence.
ACCESSION   CD352400
VERSION     CD352400.1  GI:31143911
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..741
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6856280"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GI0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector . The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      139 a    238 c    219 g    145 t
ORIGIN

  Query Match             11.8%;  Score 726.6;  DB 14;  Length 741;
  Best Local Similarity   98.8%;  Pred. No. 7e-112;
  Matches  732;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        315 GGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTGGCCCACAGCCC 374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTGGCCCACAGCCC 60

Qy        375 GCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTCGAAGACTTCTC 434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTCGAAGACTTCTC 120

Qy        435 TCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTTGAATGCCACAC 494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTTGAATGCCACAC 180

Qy        495 CAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACCGTGCCTGAGGA 554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACCGTGCCTGAGGA 240

Qy        555 GCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTCCAGGACAGTGA 614
              | |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        241 GTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATGTCCAGGACAGTGA 300

Qy        615 TGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTCAGCCAGGAGGC 674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTCAGCCAGGAGGC 360

Qy        675 CTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTCAT 734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTCAT 420

Qy        735 TGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTGATGGAGTGCGT 794
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        421 TGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTGATGGAGTGCGT 480

Qy        795 GGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATCTC 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATCTC 540

Qy        855 CACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCA 914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCA 600

Qy        915 CTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGC 974
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
```

```
Db        601 CTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGC 660

Qy        975 TGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCG 1034
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        661 TGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCG 720

Qy       1035 GACGCGGGCCAGCACCGCGCG 1055
              |||||||||||||||||||||
Db        721 GACGCGGGCCAGCACCGCGCG 741


RESULT 7
BU840589
LOCUS       BU840589                 944 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_8946353 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6490299
            5', mRNA sequence.
ACCESSION   BU840589
VERSION     BU840589.1  GI:24024989
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14039  row: o  column: 04
            High quality sequence stop: 535.
FEATURES             Location/Qualifiers
     source          1..944
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6490299"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_94"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      195 a    270 c    227 g    252 t
ORIGIN

  Query Match             11.7%;  Score 720.6;  DB 13;  Length 944;
  Best Local Similarity   96.1%;  Pred. No. 6.8e-111;
  Matches  793;  Conservative    0;  Mismatches   24;  Indels    8;  Gaps    5;

Qy       4254 TAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTA 4313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAAGGTCCAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTA 60

Qy       4314 CCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCT 4373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCTAATTTTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCT 120

Qy       4374 CTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCC 4433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTGCCCCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCC 180

Qy       4434 TGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTC 4493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGGGCCCTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTC 240

Qy       4494 TGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTT 4553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCCATGCTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTT 300

Qy       4554 CCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCAC 4613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTAACAACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCAC 360

Qy       4614 AGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAG 4673
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        361 AGCTCCCCTGCCAACCGCCTTCCTGGGCAGATGA-CCCACTCCAAGTAGATCTGGAAAAG 419

Qy       4674 ACCCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGG 4733
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 A-CCTTGTGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGG 478

Qy       4734 AAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAG 4793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AAGAGATGAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAG 538

Qy       4794 AGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACG 4853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGGCTGAAAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACG 598

Qy       4854 TGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCT 4913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        599 TGCAGTCCTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCCTGGCT 658

Qy       4914 CAGTTCCCTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAG 4973
              |||||||||||||||||| ||||||||||||||||||||||||||| |||| ||||||| 
Db        659 CAGTTCCCTGACCCTGGACCCCTGAACCTGGCTTCAGAGTAGCAAAAAGTTCCCTCCAA- 717

Qy       4974 ATGCTGTAAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTA--TTTTATCT 5031
              |||||| || ||||| |||||||||||||||||||||||||  |||| |  |||  | ||
Db        718 ATGCTGGAAGGGGAGCCTTTGCATAGGAAAAGGGCGGCTGGTCCATTTAATTTTAACCT 777

Qy       5032 TTCTTTACACTGAA---TCCCAAAATCATCTTACCACAAAGGGCC 5073
              | | |||||| |     || |||     ||||||||||||| |
Db        778 TGCCTTACACCGGAATCCCCAAAATCATCCTTACCACAAAGGGGC 822


RESULT 8
BU705518
LOCUS       BU705518                 808 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-FO0-cac-e-03-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE: 6409322 5', mRNA sequence.
ACCESSION   BU705518
VERSION     BU705518.1  GI:23635078
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
             The following repetitive elements were found in this cDNA
            sequence: 30-59, >(TAAAA)n#Simple_repeat (matched compliment)
            Seq primer: pYX-5.
```

```
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6409322"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      197 a    161 c    154 g    294 t      2 others
ORIGIN

  Query Match             11.6%;  Score 714.8;  DB 13;  Length 808;
  Best Local Similarity   97.7%;  Pred. No. 6.6e-110;
  Matches  756;  Conservative    0;  Mismatches   14;  Indels    4;  Gaps    3;

Qy       5265 CCTTTATTTATTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAA 5324
              ||||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         28 CCTTTTTTTTTTTTATTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAA 87

Qy       5325 AGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTC 5384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 AGCGTCTGTGTTTATTAGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTC 147

Qy       5385 TCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATC 5444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 TCCTCCCCTCTTCAGCCAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATC 207

Qy       5445 TCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCA 5504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 TCATTAAAGACAGGATTGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCA 267

Qy       5505 ACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGT 5564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 ACACCACCGTATGTTTCTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGT 327

Qy       5565 TAGATTCTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCAT 5624
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 TAGATTTTTTTCCTCCTTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCAT 387

Qy       5625 CACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAA 5684
              ||||||||||||||||||||||||||||||||||||||||||||  | |  |||||||||
Db        388 CACACCACATAGTCATTCCTCATGTTCTTGTTTAACAAGCACT--GAGGTTCTGGTTTAA 445

Qy       5685 ATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCT 5743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        446 ATTAAATAGCTGCAAATGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTTCT 505

Qy       5744 CAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTA 5803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 CAAAAGCAAATAAGTAATAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTA 565

Qy       5804 GCCAGAACTCACTAGCATTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTT 5863
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        566 GCCAGAACTCACTAGCATTATATATGATTTGGGTTCTGAGTAACTGGGGAGTGTTAGCTT 625

Qy       5864 TGTGACTTTGTAGCACCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAG 5923
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        626 TGTGACTTTGTAGCAGCAGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAG 685

Qy       5924 TCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTT 5983
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db        686 TCCCTTTGTCGTTTGCCATGGATGCCTTAAGTTCTTTGGAGTCTCATTTANGAATTTCTT 745

Qy       5984 TTCTCGAAGCATGACAAGTGTATCGCAATACTTACATGCTCACTCGTTTACCTG 6037
              |||||||||||||||||||||||  ||| |||||||||||||||||||||||||
Db        746 TTCTCGAAGCATGACAAGTGTAT-GCANTACTTACATGCTCACTCGTTTACCTG 798


RESULT 9
CA751384
LOCUS       CA751384               719 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  UI-M-FO0-cdn-j-21-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE: 6830806 5', mRNA sequence.
ACCESSION   CA751384
VERSION     CA751384.1  GI:25579257
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6830806"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      112 a    234 c    235 g    135 t      3 others
ORIGIN

  Query Match             11.4%;  Score 706.4;  DB 14;  Length 719;
  Best Local Similarity   98.7%;  Pred. No. 1.7e-108;
  Matches  710;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;
```

```
Qy    673 GCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTC 732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTC 60

Qy    733 ATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTGATGGAGTGC 792
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     61 ATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTGATGGAGTGC 120

Qy    793 GTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATC 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATC 180

Qy    853 TCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGG 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGG 240

Qy    913 CACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCG 972
          ||||||||||||||||||||||||||||||||||||| ||||||||| ||||||||||||
Db    241 CACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCG 300

Qy    973 GCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCG 1032
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    301 GCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCG 360

Qy   1033 CGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCG 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCG 420

Qy   1093 CTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGC 1152
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGCACTGGCTGCATGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGC 480

Qy   1153 GGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGC 1212
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    481 GGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGC 540

Qy   1213 GTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGC 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    541 GTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTANTGGTGCGC 600

Qy   1273 GAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTG 1332
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGGGGCTGCCCANCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTG 660

Qy   1333 CTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCA 1391
          ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTGGTGGCCTGNGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCA 719
```

```
RESULT 10
BU701299
LOCUS       BU701299                 740 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-EX0-bwy-i-18-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
            IMAGE: 5704793 5', mRNA sequence.
ACCESSION   BU701299
VERSION     BU701299.1  GI:23624921
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..740
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 5704793"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EX0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instutites of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      173 a    225 c    200 g    142 t
ORIGIN

  Query Match             11.4%;  Score 703.8;  DB 13;  Length 740;
  Best Local Similarity   99.3%;  Pred. No. 4.7e-108;
  Matches  738;  Conservative    0;  Mismatches    2;  Indels    3;  Gaps    3;
```

```
Qy   1959 CCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGC 2018
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      1 CCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGC 60

Qy   2019 AGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCG 2078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCG 120

Qy   2079 GCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGA 2138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 GCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGA 180

Qy   2139 GGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAA 2198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAA 240

Qy   2199 GACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGC 2258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGC 300

Qy   2259 CCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGA 2318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGA 360

Qy   2319 CTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAA 2378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAA 420

Qy   2379 TGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAA 2438
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    421 TGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGA-ATTCTCATTGGCGGCCTGAA 479

Qy   2439 ACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTT 2498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 ACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTT 539
```

```
Qy       2499 TGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCT 2558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCT 599

Qy       2559 GCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCC 2618
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        600 GCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGT-CCCCCACGGAGCC 658

Qy       2619 CAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACA 2678
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db        659 CAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAA-CACACCCAGCCCGAACA 717

Qy       2679 CCAGTGGACACTGCTCACCACAG 2701
              |||||||||||||||||||||||
Db        718 CCAGTGGACACTGCTCACCACAG 740
```

RESULT 11
CB524810

```
LOCUS       CB524810                 713 bp    mRNA    linear   EST 28-MAR-2003
DEFINITION  UI-M-FY0-cey-g-12-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6843781 5', mRNA sequence.
ACCESSION   CB524810
VERSION     CB524810.1  GI:29358165
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..713
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6843781"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      141 a    235 c    185 g    148 t      4 others
ORIGIN
```

```
  Query Match             11.2%;  Score 690;  DB 14;  Length 713;
  Best Local Similarity   98.6%;  Pred. No. 9.7e-106;
  Matches  704;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy       2489 ATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTC 2548
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          1 ATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTC-ACACCTC 59

Qy       2549 CTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCC 2608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 CTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCC 119

Qy       2609 CCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCC 2668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 CCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCC 179

Qy       2669 AGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCC 2728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 AGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCC 239

Qy       2729 ATGGCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGC 2788
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 ATGGCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGC 299

Qy       2789 CTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCT 2848
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        300 CTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAAAGACATTCTCAGACTCCT 359

Qy       2849 TGGATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCC 2908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 TGGATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCC 419

Qy       2909 TGGCCTGCATGTGTGCTGGCCTACGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGT 2968
              ||||||||||||||||||||||||||| |||||||||||||||||||| |||||||||||
Db        420 TGGCCTGCATGTGTGCTGGCCTACGACGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGT 479

Qy       2969 CCTCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCA 3028
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db        480 CCTCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCA 539

Qy       3029 GTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCC 3088
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        540 GTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCNCCAGGATTGGTCCC 599

Qy       3089 CACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTT 3148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 CACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTT 659

Qy       3149 CAGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGAC 3202
              ||||||||||||||||||||| ||||||||||||||| ||||| ||||||||||
Db        660 CAGATTGCCGGGGCCACTCCANGAGAAAGATCTCCTGNGCTCANGCAGGGGGAC 713
```

RESULT 12
CD351162

```
LOCUS       CD351162                 699 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-FY0-cft-p-13-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6852446 5', mRNA sequence.
ACCESSION   CD351162
VERSION     CD351162.1  GI:31142737
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
```

```
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..699
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6852446"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      135 a    221 c    202 g    139 t      2 others
ORIGIN

  Query Match             11.1%;  Score 685.8;  DB 14;  Length 699;
  Best Local Similarity   98.7%;  Pred. No. 4.9e-105;
  Matches  690;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2753 ATTTCTTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGG 2812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTTCTTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGG 60

Qy       2813 ATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCA 2872
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGTGATTACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCA 120

Qy       2873 TCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTAC 2932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTAC 180

Qy       2933 GACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACC 2992
              ||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        181 GACGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACC 240

Qy       2993 CAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGG 3052
              |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db        241 CAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGG 300

Qy       3053 AGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACA 3112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACA 360

Qy       3113 AGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGA 3172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGA 420

Qy       3173 GAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTG 3232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTG 480

Qy       3233 AGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGG 3292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGG 540

Qy       3293 GAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGA 3352
              |||||||||||||||||||||||||||||||||| ||||||| |||| ||||||||||||
Db        541 GAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATANAGAGCAACGGGA 600

Qy       3353 GAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCG 3412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCG 660

Qy       3413 ACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCC 3451
              ||||||||||||||| |||||||||||||||||||||||
Db        661 ACTTCGGTGCATCCANAGGATGTCCTGACCTCCACCTCC 699


RESULT 13
BI185886
LOCUS       BI185886                 900 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  UNL-P-FN-df-h-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-df-h-02-0-UNL 3', mRNA sequence.
ACCESSION   BI185886
VERSION     BI185886.1  GI:14660295
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Mamm. Genome 14 (1), 65-70 (2003)
  MEDLINE   22419904
   PUBMED   12532269
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..900
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-df-h-02-0-UNL"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UNL-P-FN"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      149 a    304 c    291 g    150 t      6 others
ORIGIN
```

```
Query Match             10.4%;  Score 639.6;  DB 12;  Length 900;
Best Local Similarity   86.1%;  Pred. No. 2.7e-97;
Matches  719;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        739 GCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTGATGGAGTGCGTGGCC 798
Db          7 GCTGCCCCGGAGAACACCACAGTGGTGTCCGGCCAGAGCGTGGTGATGGAGTGTGTGGCC 66

Qy        799 TCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATCTCCACG 858
Db         67 TCGGCTGACCCCACCCCTATCGTGTCCTGGGTCCGACAGGACGGGAAGCCTATTTCTACC 126

Qy        859 GATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCT 918
Db        127 GATGTCATCGTCCTGGGCCGCACCAACCTGCTAATCACCAGCGCGCANCC-CGGCACTCT 185

Qy        919 GGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCT 978
Db        186 GGCGTCTACGTCTGCAGAGCCAACAAGCCCCGTACGCGCGACTTTGCCACCGCCGCCGCC 245

Qy        979 GAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACG 1038
Db        246 GAGCTCCGCGTGCTGGCGGCTCCTGCCATCTCGCAGGCGCCCGAGGCGCTATCTCGGACG 305

Qy       1039 CGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCAC 1098
Db        306 AGGGCGAGCACCGCGCGCTTCGTGTGCCGCGCAACGGGTGAGCCGCGGCCCGCGTTGCGG 365

Qy       1099 TGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGC 1158
Db        366 TGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGTGTCAAGGTGCAGGGCGGCGGC 425

Qy       1159 GGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCA 1218
Db        426 AGCAGCCTGGTCATCACGCAGATCGGCCTGCAGGATGCCGGCTACTACCAGTGCGTGGCC 485

Qy       1219 GAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGG 1278
Db        486 GAGAACAACGCGGGCACGGCGTGCGCCGCCGCGCCGCTGGCAGTAGTAGTGCGCGAGGGT 545

Qy       1279 CTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTG 1338
Db        546 CTGCCCAGCGCCCCCACGCGGGTCACCGCCACGCCCCTGAGCAGCTCCACTGTGCTGGTG 605

Qy       1339 GCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAA 1398
Db        606 GCCTGGGAGCGGCCCGAGCTGCACAGCGAGCAGATCATCGGTTTCTCCCTTCACTACCAG 665

Qy       1399 AAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAG 1458
Db        666 AAGGCACGGGGCATGGACAGTGTGGAATACCAGTTTGCCGTGAACAATGACACCACTGAG 725

Qy       1459 CTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCC 1518
Db        726 CTACAGGTTCGGGACCTGGAGCCCAACACAGACTATGAGTTCTATGTGGTGGCCTACTCA 785

Qy       1519 CAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATG 1573
Db        786 CAGCTGGGGGCCAGCCGCACCTCCACCCCAGCGCTGGTCCACACATTGGATGATG 840
```

RESULT 14
CB518504

```
LOCUS       CB518504               631 bp    mRNA    linear   EST 28-MAR-2003
DEFINITION  UI-M-GH0-cec-h-04-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
            IMAGE: 6837653 5', mRNA sequence.
ACCESSION   CB518504
VERSION     CB518504.1  GI:29351859
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6837653"
                     /tissue_type="Whole brain"
                     /dev_stage="1, 5, and 15 days newborn"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GH0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CGAACTGAAT. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      151 a    183 c    180 g    117 t
ORIGIN

  Query Match             10.2%;  Score 629.4;  DB 14;  Length 631;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-95;
  Matches  630;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1406 GGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGG 1465
Db          1 GGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGG 60

Qy       1466 TTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGG 1525
Db         61 TTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGG 120

Qy       1526 GGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAG 1585
Db        121 GGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAG 180

Qy       1586 CACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCC 1645
Db        181 CACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCC 240

Qy       1646 TGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGA 1705
Db        241 TGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGA 300

Qy       1706 AGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACT 1765
Db        301 AGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACT 360

Qy       1766 CACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATG 1825
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 CACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATG 420

Qy        1826 GAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTC 1885
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTC 480

Qy        1886 CCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGC 1945
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGC 540

Qy        1946 AGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAA 2005
               ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         541 AGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAA 600

Qy        2006 CAGAGGAGGAGGCAGATGGTGACCGCCCCCC 2036
               |||||||||||||||||||||||||||||||
Db         601 CAGAGGAGGAGGCAGATGGTGACCGCCCCCC 631
```

```
RESULT 15
BY719391
LOCUS       BY719391                 671 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY719391 RIKEN full-length enriched, 12 days embryo male wolffian
            duct Mus musculus cDNA clone 6720465A10 5', mRNA sequence.
ACCESSION   BY719391
VERSION     BY719391.1  GI:27132508
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   MEDLINE  22354683
    PUBMED  12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..671
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6720465A10"
                     /sex="male"
                     /tissue_type="wolffian duct includes surrounding region"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     male wolffian duct"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT      183 a    132 c    119 g    237 t
ORIGIN

  Query Match             10.0%;  Score 618.2;  DB 14;  Length 671;
  Best Local Similarity   96.9%;  Pred. No. 1.1e-93;
  Matches  652;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    2;

Qy        5233 TGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTATTTTATTTTATT 5292
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 TGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTATTTTATTTTATT 60

Qy        5293 TTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGT 5352
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATTAGCCTTGTGTGT 120

Qy        5353 CACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATGATA 5412
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGCCAGCCTATGATA 180
```

```
Qy       5413 ACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAAC 5472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGATTGTCACTTGAAC 240

Qy       5473 TACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGCTCT 5532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTTCTTCATTGCTCT 300

Qy       5533 GAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCCTTGCCTTGAATG 5592
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        301 GAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTTTTTTCCTCCTTGCCTTGAATG 360

Qy       5593 AAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCATGTTCT 5652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCATGTTCT 420

Qy       5653 TGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGACAATTTA 5712
              ||||||||||||||     | |    ||||||||||||||||||||||||||||||||||
Db        421 TGTTTAACAAGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGACAATTTA 478

Qy       5713 TAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAATCTGGTA 5771
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        479 TAACCCATTAGGCTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAATCTGGTA 538

Qy       5772 TCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATATATGAT 5831
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        539 TCTGCCTATAACTCACAGTTGATAAGAAAATAGCCAGAACTCACTAGCATTATATATGAT 598

Qy       5832 TGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTCTTATTA 5891
              | ||||||||||||||||||||||||||||||||||| |||| | || ||||||||||||
Db        599 TTGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGAATTTGGAACAACAGGTCTTATTA 658

Qy       5892 GGAAAGTCTGTTG 5904
               ||||   || ||
Db        659 AGAAAAGTTGGTG 671
```

Search completed: October 15, 2003, 11:38:58
Job time : 14070.8 secs

```
                     GenCore version 5.1.6
              Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         October 14, 2003, 06:22:55 ; Search time 1163.58 Seconds
                                             (without alignments)
                                             14327.989 Million cell updates/sec

Title:          US-09-754-997A-1
Perfect score:  6176
Sequence:       1 atggcgcgggcggacacggg..........taataaactattttccagac 6176

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                 25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		6176	100.0	6176	22	AAD10021	Mouse Nope (neighb
2		2804.2	45.4	3753	22	AAI67204	Nucleotide sequenc
3		2796	45.3	2796	22	AAD10022	Mouse Nope (neighb
4		2681.8	43.4	3741	24	ABK11101	DNA encoding human
5		2118.8	34.3	3450	24	ABZ11219	Human polynucleoti
6		1522.4	24.7	2223	22	AAI67203	Nucleotide sequenc
7		825	13.4	825	22	AAD10023	Mouse Nope (neighb
8		337.2	5.5	680	21	AAC79893	Human secreted pro
9		316.6	5.1	2976	25	ABX71437	Human cDNA encodin
10		316.6	5.1	3210	25	ABX71432	Human cDNA encodin
11		316.6	5.1	3219	25	ABX71440	Human cDNA encodin
12		316.6	5.1	3874	25	ABX71447	Human cDNA encodin
13		316.4	5.1	2700	24	AAD36777	Human immunoglobul
14		315.4	5.1	3453	25	ABX71435	Human cDNA encodin
15		289	4.7	2695	23	AAS76339	DNA encoding novel
16		288	4.7	288	22	AAD10030	Mouse Nope (neighb
17		252	4.1	252	22	AAD10032	Mouse Nope (neighb
18		249	4.0	249	22	AAD10028	Mouse Nope (neighb
19		249	4.0	249	22	AAD10029	Mouse Nope (neighb
20		246	4.0	246	22	AAD10031	Mouse Nope (neighb
21		243	3.9	243	22	AAD10024	Mouse Nope (neighb
22		217	3.5	2481	25	ABX71439	Human cDNA encodin
23		217	3.5	2715	25	ABX71434	Human cDNA encodin
24		217	3.5	2724	25	ABX71441	Human cDNA encodin
25		217	3.5	2958	25	ABX71436	Human cDNA encodin
26		213.8	3.5	2139	25	ABX71442	Human cDNA encodin
27		213.8	3.5	2382	25	ABX71445	Human cDNA encodin
28		201.8	3.3	563	24	ABQ25578	Oligonucleotide fo
29	c	201.8	3.3	563	24	ABQ25579	Oligonucleotide fo
30		200.2	3.2	3330	23	AAS76658	DNA encoding novel
31		195	3.2	195	22	AAD10027	Mouse Nope (neighb
32		192	3.1	192	22	AAD10025	Mouse Nope (neighb
33	c	190.4	3.1	563	24	ABQ25580	Oligonucleotide fo
34		190.4	3.1	563	24	ABQ25581	Oligonucleotide fo
35		189	3.1	189	22	AAD10026	Mouse Nope (neighb
36		151.2	2.4	1875	25	ABX71443	Human cDNA encodin
37		119	1.9	909	25	ABX71438	Human cDNA encodin
38		119	1.9	1143	25	ABX71433	Human cDNA encodin
39		117	1.9	1644	25	ABX71444	Human cDNA encodin
40		117	1.9	1887	25	ABX71446	Human cDNA encodin
41		99.4	1.6	2450	23	AAS76657	DNA encoding novel
42		93.8	1.5	2560	25	ABX71188	Novel human cDNA s
43		85.6	1.4	439	24	ABS69793	Novel murine polyn
44		75	1.2	370	22	AAS33787	Human cDNA encodin
45		71	1.1	4277	21	AAA37039	Human PRO1190 (UNQ

ALIGNMENTS

```
RESULT 1
AAD10021
ID   AAD10021 standard; cDNA; 6176 BP.
XX
AC   AAD10021;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3759
FT                   /*tag=  a
FT                   /product= "Mouse Nope protein"
FT   sig_peptide     1..63
FT                   /*tag=  b
FT   mat_peptide     64..3756
FT                   /*tag=  c
FT                   /product= "Mature mouse Nope protein"
FT   misc_feature    1..3756
FT                   /*tag=  d
FT                   /note= "This region is specifically referred as
```

```
FT                   SEQ ID NO: 45 in the specification"
FT   misc_feature    64..2859
FT                   /*tag=  e
FT                   /note= "Extracellular domain"
FT   misc_feature    139..381
FT                   /*tag=  f
FT                   /label= Ig1
FT                   /note= "Immunoglobulin domain 1"
FT   misc_feature    463..654
FT                   /*tag=  g
FT                   /label= Ig2
FT                   /note= "Immunoglobulin domain 2"
FT   misc_feature    766..954
FT                   /*tag=  h
FT                   /label= Ig3
FT                   /note= "Immunoglobulin domain 3"
FT   misc_feature    1039..1233
FT                   /*tag=  i
FT                   /label= Ig4
FT                   /note= "Immunoglobulin domain 4"
FT   misc_feature    1285..1533
FT                   /*tag=  j
FT                   /label= Fn1
FT                   /note= "Fibronectin-type III domain 1"
FT   misc_feature    1579..1827
FT                   /*tag=  k
FT                   /label= Fn2
FT                   /note= "Fibronectin-type III domain 2"
FT   misc_feature    1888..2175
FT                   /*tag=  l
FT                   /label= Fn3
FT                   /note= "Fibronectin-type III domain 3"
FT   misc_feature    2248..2493
FT                   /*tag=  m
FT                   /label= Fn4
FT                   /note= "Fibronectin-type III domain 4"
FT   misc_feature    2542..2793
FT                   /*tag=  n
FT                   /label= Fn5
FT                   /note= "Fibronectin-type III domain 5"
FT   misc_feature    2932..3756
FT                   /*tag=  o
FT                   /note= "Intracellular domain"
FT   3'UTR           3819..5682
FT                   /*tag=  p
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAE05251.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 10; Fig 2A; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) cDNA.
XX
SQ   Sequence 6176 BP; 1316 A; 1756 C; 1634 G; 1470 T; 0 other;

  Query Match             100.0%;  Score 6176;  DB 22;  Length 6176;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
```

```
Qy       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780

Qy       3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840

Qy       3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900

Qy       3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960

Qy       3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020

Qy       4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080

Qy       4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140

Qy       4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200

Qy       4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260

Qy       4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320

Qy       4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380

Qy       4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440

Qy       4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500

Qy       4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560

Qy       4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620

Qy       4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680

Qy       4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740

Qy       4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800

Qy       4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860

Qy       4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920

Qy       4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980

Qy       4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040

Qy       5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100

Qy       5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160

Qy       5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220

Qy         5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280

Qy         5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340

Qy         5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400

Qy         5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460

Qy         5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520

Qy         5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580

Qy         5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640

Qy         5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700

Qy         5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760

Qy         5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820

Qy         5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880

Qy         5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940

Qy         5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000

Qy         6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060

Qy         6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120

Qy         6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
```

RESULT 2
AAI67204
```
ID   AAI67204 standard; DNA; 3753 BP.
XX
AC   AAI67204;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Nucleotide sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdoch PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   P-PSDB; AAG65914.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer  -
XX
PS   Claim 2; Page 55-56; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence 3753 BP; 714 A; 1220 C; 1152 G; 667 T; 0 other;

  Query Match             45.4%;  Score 2804.2;  DB 22;  Length 3753;
  Best Local Similarity   84.7%;  Pred. No. 0;
  Matches 3171;  Conservative    0;  Mismatches  568;  Indels    6;  Gaps    2;

Qy          3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62
              || |   || | |   |||||||||||||| | | |||||||||||||| || |||||||
Db          9 GGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 68

Qy         63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122
              |||||||||||  |||||||||||||| ||||| | |||||||||||| | ||| || ||
Db         69 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACT 128

Qy        123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182
              |||||||||||||||||| ||||||||||  |||||  |||| |     |||||||| | ||
Db        129 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 188
```

```
Qy    183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242
Db    189 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 248

Qy    243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302
Db    249 CCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 308

Qy    303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362
Db    309 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCT 365

Qy    363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422
Db    366 AGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACT 425

Qy    423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482
Db    426 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 485

Qy    483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542
Db    486 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 545

Qy    543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602
Db    546 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 605

Qy    603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662
Db    606 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 665

Qy    663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722
Db    666 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCA 725

Qy    723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782
Db    726 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 785

Qy    783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842
Db    786 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGG 845

Qy    843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902
Db    846 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 905

Qy    903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962
Db    906 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 965

Qy    963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGA 1022
Db    966 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGA 1025

Qy   1023 GGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCC 1082
Db   1026 GGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCC 1085

Qy   1083 ACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAA 1142
Db   1086 GCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAA 1145

Qy   1143 GGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTA 1202
Db   1146 GGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTA 1205

Qy   1203 CTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGT 1262
Db   1206 CTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGT 1265

Qy   1263 AGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAG 1322
Db   1266 GGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAG 1325

Qy   1323 CTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTT 1382
Db   1326 CTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTT 1385

Qy   1383 CTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAA 1442
Db   1386 CTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAA 1445

Qy   1443 CAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTA 1502
Db   1446 CAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTA 1505

Qy   1503 CGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATAC 1562
Db   1506 CGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACAC 1565

Qy   1563 ACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGA 1622
Db   1566 ACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGA 1625

Qy   1623 CATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTA 1682
Db   1626 CATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTA 1685

Qy   1683 CAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAA 1742
Db   1686 CAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAA 1745

Qy   1743 TGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTC 1802
Db   1746 TGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTC 1805

Qy   1803 AGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGG 1862
Db   1806 GGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAG 1865

Qy   1863 TGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGAT 1922
Db   1866 TATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGAT 1925

Qy   1923 GGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAA 1982
Db   1926 GGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAA 1985

Qy   1983 ACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGG 2042
Db   1986 ACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGG 2045

Qy   2043 TCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTA 2102
Db   2046 CCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTA 2105

Qy   2103 TGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAA 2162
Db   2106 TGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAA 2165

Qy   2163 ACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGA 2222
Db   2166 ACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGA 2225

Qy   2223 CCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAG 2282
Db   2226 CATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAG 2285

Qy   2283 CTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAA 2342
Db   2286 CTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAA 2345

Qy   2343 CTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATAC 2402
```

```
Db   2346 CTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACAC 2405

Qy   2403 CAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGC 2462
Db   2406 CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGC 2465

Qy   2463 GGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCAC 2522
Db   2466 AGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCAC 2525

Qy   2523 CCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATC 2582
Db   2526 CCTGCCTGACCGGCCCTCCACACCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTC 2585

Qy   2583 CACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCT 2642
Db   2586 CACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCT 2645

Qy   2643 AATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGA 2702
Db   2646 GATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCA 2705

Qy   2703 GGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAA 2762
Db   2706 GGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAA 2765

Qy   2763 GATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTAC 2822
Db   2766 GATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCAC 2825

Qy   2823 TCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTGGG 2882
Db   2826 GCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATCATCGTGGG 2885

Qy   2883 TGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGCTC 2942
Db   2886 TGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCC 2945

Qy   2943 CCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTCTA 3002
Db   2946 CCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTA 3005

Qy   3003 CACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGT 3062
Db   3006 CTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGT 3065

Qy   3063 GCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGT 3122
Db   3066 GCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGT 3125

Qy   3123 ACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTC 3182
Db   3126 GCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGATCTC 3185

Qy   3183 CTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCA 3242
Db   3186 CTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCA 3245

Qy   3243 GGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCA 3302
Db   3246 GG---CAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCA 3302

Qy   3303 GACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAGCC 3362
Db   3303 GACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGAGGAAGAAGTC 3362

Qy   3363 GTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGC 3422
Db   3363 ACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGC 3422

Qy   3423 ATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGC 3482
Db   3423 ATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCC 3482

Qy   3483 AGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGG 3542
Db   3483 AGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGA 3542

Qy   3543 CAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCC 3602
Db   3543 CAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCC 3602

Qy   3603 AGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGA 3662
Db   3603 AGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGA 3662

Qy   3663 GGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGCCT 3722
Db   3663 GACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCT 3722

Qy   3723 TCCCAGGGCCCCTGTCTCCTCTGCT 3747
Db   3723 GCCCAGATCCCCGGTCTCCTCCTCT 3747
```

RESULT 3
AAD10022

```
ID   AAD10022 standard; cDNA; 2796 BP.
XX
AC   AAD10022;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) extracellular domain cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; extracellular domain; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2796
FT                   /*tag=  a
FT                   /product= "Mouse Nope extracellular domain"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAE05252.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 11; Page 77-81; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
```

```
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) extracellular domain cDNA.
XX
SQ   Sequence 2796 BP; 591 A; 849 C; 821 G; 535 T; 0 other;

  Query Match             45.3%;  Score 2796;  DB 22;  Length 2796;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
```

```
Qy       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy       2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy       2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy       2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy       2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy       2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy       2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy       2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy       2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy       2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy       2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy       2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859
              ||||||||||||||||||||||||||||||||||||
Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

```
RESULT 4
ABK11101
ID   ABK11101 standard; DNA; 3741 BP.
XX
AC   ABK11101;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding human NOV1 protein, homologue of NOPE/PUNC Ig proteins.
XX
KW   Human; NOVX-associated disorder; developmental disorder; blood disorder;
KW   endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
KW   respiratory disorder; inflammatory disorder; reproductive disorder;
KW   neurodegenerative disorder; autoimmune disorder; infectious disease;
KW   cardiovascular disorder; cell signal processing; PUNC immunoglobulin;
KW   metabolic pathway modulation; NOPE; Ig; NOV1; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3741
FT                   /*tag=  a
FT                   /partial
FT                   /product= "NOV1"
FT                   /note= "This sequence lacks a stop codon"
XX
PN   WO200206329-A2.
XX
PD   24-JAN-2002.
XX
PF   18-JUL-2001; 2001WO-US22709.
XX
PR   18-JUL-2000; 2000US-218870P.
PR   18-JUL-2000; 2000US-218875P.
PR   18-JUL-2000; 2000US-218901P.
PR   24-JUL-2000; 2000US-220273P.
PR   26-JUL-2000; 2000US-220912P.
PR   27-JUL-2000; 2000US-221233P.
PR   28-JUL-2000; 2000US-221650P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Rastelli L,  Shimkets RA,  Zerhusen B,  Malyankar UM,  Padigaru M;
XX
DR   WPI; 2002-179781/23.
DR   P-PSDB; AAU77405.
XX
PT   Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT   polypeptides, useful for treating developmental disorders, endocrine
PT   disorders, vascular disorders, infectious diseases and
PT   neurodegenerative disorders -
XX
PS   Claim 9; Page 8-9; 178pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5A,
CC   NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC   The NOVX polypeptides are related to NOPE, cadherin, interferon
CC   alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin
CC   polypeptides. The sequences of the invention are useful for identifying
CC   an agent (a cellular receptor or downstream effector) that binds to
CC   the NOVX polypeptide, or an agent that modulates it's expression or
CC   activity. They are useful for treating or preventing NOVX-associated
CC   disorders such as developmental disorders, endocrine disorders, vascular
CC   diseases, gastrointestinal disorders, respiratory disorders, inflammatory
CC   disorders, blood disorders, reproductive disorders, neurodegenerative
CC   disorders, autoimmune and immune disorders, infectious diseases,
CC   cardiovascular disorders, cancers, and other disorders related to cell
CC   signal processing and metabolic pathway modulation. The present sequence
CC   encodes the human NOV1 protein.
XX
SQ   Sequence 3741 BP; 708 A; 1209 C; 1150 G; 673 T; 1 other;

  Query Match             43.4%;  Score 2681.8;  DB 24;  Length 3741;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 3141;  Conservative    0;  Mismatches  583;  Indels   30;  Gaps    8;

Qy          3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62
                || |   || | |   ||||||||||||| | | |||||||||||||| || |||||||
Db          9 GGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 68
```

```
Qy       63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122
Db       69 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACT 128

Qy      123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182
Db      129 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 188

Qy      183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242
Db      189 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 248

Qy      243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302
Db      249 CCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 308

Qy      303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362
Db      309 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCT 365

Qy      363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422
Db      366 AGCCCACGGNCCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCAGTCT 425

Qy      423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482
Db      426 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 485

Qy      483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542
Db      486 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 545

Qy      543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602
Db      546 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 605

Qy      603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662
Db      606 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 665

Qy      663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722
Db      666 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACCAGGGGGCA 725

Qy      723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782
Db      726 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 785

Qy      783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842
Db      786 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA---GACGG 842

Qy      843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902
Db      843 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 902

Qy      903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962
Db      903 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 962

Qy      963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTT---GCTGCCCCAGCCATCTCGCAGGCACC 1019
Db      963 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACTCAGGCGCC 1022

Qy     1020 CGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGA 1079
Db     1023 CGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGA 1082

Qy     1080 GCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGT 1139
Db     1083 GCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGT 1142

Qy     1140 CAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGG 1199
Db     1143 CAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGG 1202

Qy     1200 CTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGC 1259
Db     1203 CTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGC 1262

Qy     1260 GGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAG 1319
Db     1263 CGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAG 1322

Qy     1320 CAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGG 1379
Db     1323 CAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGG 1382

Qy     1380 CTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGT 1439
Db     1383 CTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAGTTTGCAGT 1442

Qy     1440 AAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTT 1499
Db     1443 GAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTT 1502

Qy     1500 CTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCA 1559
Db     1503 CTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCA 1562

Qy     1560 TACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTC 1619
Db     1563 CACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTC 1622

Qy     1620 GGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAA 1679
Db     1623 GGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAA 1682

Qy     1680 GTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGG 1739
Db     1683 GTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGG 1742

Qy     1740 AAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGAT 1799
Db     1743 AAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGAT 1802

Qy     1800 TTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACC 1859
Db     1803 TTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCC 1862

Qy     1860 TGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAA 1919
Db     1863 CAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAA 1922

Qy     1920 GATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATA 1979
Db     1923 GATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTA 1982

Qy     1980 CAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGG 2039
Db     1983 CAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGG 2042

Qy     2040 GGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCA 2099
Db     2043 GGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCA 2102

Qy     2100 GTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAA 2159
Db     2103 GTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAA 2162

Qy     2160 CAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCC 2219
Db     2163 CAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACC 2222

Qy     2220 AGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAA 2279
```

```
Db   2223 AGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAA 2282

Qy   2280 CAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGT 2339
Db   2283 CAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGT 2342

Qy   2340 CAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTA 2399
Db   2343 CAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTA 2402

Qy   2400 TACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTT 2459
Db   2403 ---CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTT 2459

Qy   2460 TGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTC 2519
Db   2460 TGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTC 2519

Qy   2520 CACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACC 2579
Db   2520 CACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACC 2579

Qy   2580 ATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTA 2639
Db   2580 GTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTA 2639

Qy   2640 TCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCAC 2699
Db   2640 TCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCAC 2699

Qy   2700 ------AGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2700 GCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTA 2759

Qy   2754 TTTCTTCAAGATGGAGGCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2760 CTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGA 2819

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCAT 2873
Db   2820 TGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCAT 2879

Qy   2874 CATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACG 2933
Db   2880 CATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCG 2939

Qy   2934 ACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCC 2993
Db   2940 CCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCC 2999

Qy   2994 AGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGA 3053
Db   3000 CGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGA 3059

Qy   3054 GTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAA 3113
Db   3060 GTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAG 3119

Qy   3114 GGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAG 3173
Db   3120 GGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAG 3179

Qy   3174 AAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGA 3233
Db   3180 AAAGGT---AAGTGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGA 3236

Qy   3234 GCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGG 3293
Db   3237 GCTGCCCCAGG---CAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGG 3293

Qy   3294 AACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAG 3353
Db   3294 AACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGATC------AGGGCAATGGGAG 3347

Qy   3354 AAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGA 3413
Db   3348 GAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGA 3407

Qy   3414 CTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACC 3473
Db   3408 CTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCC 3467

Qy   3474 ACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGA 3533
Db   3468 CCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGC 3527

Qy   3534 CTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCAC 3593
Db   3528 CTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTAC 3587

Qy   3594 CTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGC 3653
Db   3588 CTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGT 3647

Qy   3654 TATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAG 3713
Db   3648 GCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAG 3707

Qy   3714 CCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCT 3747
Db   3708 CCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741
```

```
RESULT 5
ABZ11219
ID   ABZ11219 standard; cDNA; 3450 BP.
XX
AC   ABZ11219;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 101.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69002.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
```

```
PS   Claim 1; SEQ ID NO 101; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3450 BP; 681 A; 1050 C; 1067 G; 652 T; 0 other;

  Query Match             34.3%;  Score 2118.8;  DB 24;  Length 3450;
  Best Local Similarity   80.8%;  Pred. No. 0;
  Matches 2599;  Conservative    0;  Mismatches  567;  Indels   50;  Gaps    9;

Qy        991 CTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACC 1050
Db        184 CTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACA 243

Qy       1051 GCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGAC 1110
Db        244 GCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAAC 303

Qy       1111 GGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTC 1170
Db        304 GGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTC 363

Qy       1171 ATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCG 1230
Db        364 ATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCG 423

Qy       1231 GGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCC 1290
Db        424 GGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCC 483

Qy       1291 CCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGG 1350
Db        484 CCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGG 543

Qy       1351 CCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGA 1410
Db        544 CCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGC 603

Qy       1411 GTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGG 1470
Db        604 ATGGACAATGTGGAATACCAGTTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGG 663

Qy       1471 GACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCC 1530
Db        664 GACCTGGAACCCAACACAGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCC 723

Qy       1531 AGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCC 1590
Db        724 AGCCGCACCTCCACCCCAGCACTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCC 783

Qy       1591 CAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCC 1650
Db        784 CAGCTCTCCCTGTCCAGCCCCAACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCC 843

Qy       1651 TCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAA 1710
Db        844 CCCAGCCTGAGCAATGGGCAGGTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAA 903

Qy       1711 GATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTT 1770
Db        904 GATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTT 963

Qy       1771 CAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTC 1830
Db        964 CAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCC 1023

Qy       1831 CCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTT 1890
Db       1024 CCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTT 1083

Qy       1891 GCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCG 1950
Db       1084 GCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCA 1143

Qy       1951 CCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAG 2010
Db       1144 CCCCCTCACCCCACCCAGATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAG 1203

Qy       2011 GAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGG 2070
Db       1204 GAGGAGGCCAATGGCGATCGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGG 1263

Qy       2071 CCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGG 2130
Db       1264 CCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGG 1323

Qy       2131 CCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGG 2190
Db       1324 CTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGG 1383

Qy       2191 AAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTG 2250
Db       1384 AAGGGCAAGACGGAGAAGGCGCCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTG 1443

Qy       2251 CCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAG 2310
Db       1444 CCTCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAA 1503

Qy       2311 AAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGG 2370
Db       1504 AAGCCAGATTTCACCACAGTCAAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGG 1563

Qy       2371 CTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGC 2430
Db       1564 CTCAGGAATGCCTCCCTGGTCACCTATTACACCAGTTCTGGAGAAGACATCCTCATTGGC 1623

Qy       2431 GGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGAT 2490
Db       1624 GGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGAT 1683

Qy       2491 GGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCT 2550
Db       1684 GGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCA 1743

Qy       2551 TCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCC 2610
Db       1744 TCCGACCTGCGACTGAGCCCCCTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCC 1803

Qy       2611 ACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAG 2670
Db       1804 ACAGAGCCCAACGGGGAGATCGTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAG 1863

Qy       2671 CCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCAT 2730
Db       1864 CCTGAGCACCAGTGGACCTTGCTCACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCAT 1923

Qy       2731 GGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCT 2790
Db       1924 GGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCT 1983
```

```
Qy       2791 GGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTG 2850
Db       1984 GGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTG 2043

Qy       2851 GATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTG 2910
Db       2044 GACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTG 2103

Qy       2911 GCCTGCATGTGTGCTGGCCTACGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCC 2970
Db       2104 GCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCC 2163

Qy       2971 TCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGT 3030
Db       2164 TCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGC 2223

Qy       3031 GTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCA 3090
Db       2224 CCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCG 2283

Qy       3091 CCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCA 3150
Db       2284 CCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCT 2343

Qy       3151 GATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGG 3210
Db       2344 GAAGGCCGGAGTCACTCCAAAAGAAAGATCTCCTGGGCTCAACCAAGCGGGCTGAGCTGG 2403

Qy       3211 GCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACC 3270
Db       2404 GCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAGG---CAGGCCCCCGGCCGGCTCTGACC 2460

Qy       3271 CGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTAT 3330
Db       2461 CGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGCTCTGGTGTAC 2520

Qy       3331 GACGGCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAA 3390
Db       2521 GACGCCATAAAGGGCAATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAG 2580

Qy       3391 GCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTC 3450
Db       2581 GCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTC 2640

Qy       3451 CAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCT 3510
Db       2641 CAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTT 2700

Qy       3511 GTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACA 3570
Db       2701 GGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCA 2760

Qy       3571 ACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGC 3630
Db       2761 GCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTAC 2820

Qy       3631 TCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAA 3690
Db       2821 CCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAA 2880

Qy       3691 GCCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAG 3750
Db       2881 TCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCTGCC 2940

Qy       3751 GTCCCCTGAGCAGA----------------AGGCAGATATGGCTCAGGAA---------- 3784
Db       2941 TAGCTCTTCCCAGAGGATGTGGTTTGGGGCAGGCAGGTATGGATCACATAGGATGCGATA 3000

Qy       3785 ---CATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGCC 3841
Db       3001 CCTGTGGCCGTGTATGTCCACATGTGTGCCTGTAGATACATCATCAAGCCCTTTGGAGCT 3060

Qy       3842 TCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAA------CTTTACTCTCCCTCATATTCTG 3895
Db       3061 TCCTAAGTTGCTTTGGCTGAGGGGAGAGGAAAACATGGATTATTCACTCCCCCATACTC 3120

Qy       3896 CATCACATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTG 3955
Db       3121 TTTGTGATACACATGTGACATGTGAAAGACATACGAGACATAGCTACATGTGATGTGCAC 3180

Qy       3956 TGTGTGTGTGTGTGTGTGTG-----TGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGC 4010
Db       3181 ATGTGTGAAGTGCATGTATGCGTACTGGTTGTTGAGCTGGGAAACCGTGGCCCAGGCAGT 3240

Qy       4011 ACTCACTGTGGCCTAGTTGACCCTCCGTGGCAGGATGGTG--TAACAGTGATCAGTGCCA 4068
Db       3241 GGTCACTACAGCCTGATTGGTCCTCCAGGTCAGAACGGTGCCCCACAGTGGTCAGTCCCC 3300

Qy       4069 GCTCTTTGAGCTTTTAGCCT--TGTCACCTAGCCTTTTATT--ACACTCTGAGAGTGTCT 4124
Db       3301 AGCCCTGTGGGCCCCCACCTCCATTCGCCCAGCCTTTTATTACACACTCTGAGAGTGTCT 3360

Qy       4125 CCAGTG-CTGTGTCTACAAAGACAGCGCCCAGCCCT 4159
Db       3361 CCAATGCCTGTCTTGACAAAGACAGTTCCCAGCCCT 3396
```

RESULT 6
AAI67203
ID   AAI67203 standard; DNA; 2223 BP.
XX
AC   AAI67203;
XX
DT   11-FEB-2002 (first entry)
XX
DE   Nucleotide sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   P-PSDB; AAG65913.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer -
XX
PS   Claim 2; Page 54-55; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant

```
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence 2223 BP; 440 A; 660 C; 735 G; 388 T; 0 other;

  Query Match             24.7%;  Score 1522.4;  DB 22;  Length 2223;
  Best Local Similarity   82.8%;  Pred. No. 0;
  Matches 1786;  Conservative    0;  Mismatches  321;  Indels   51;  Gaps    2;

Qy         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123
Db         76 GGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACTG 135

Qy        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183
Db        136 CAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGCC 195

Qy        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243
Db        196 GCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGAC 255

Qy        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303
Db        256 CACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGGC 315

Qy        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363
Db        316 AG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCTA 372

Qy        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423
Db        373 GCCCACGGCCCCCCTGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACTC 432

Qy        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483
Db        433 GCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTTT 492

Qy        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543
Db        493 GAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGACA 552

Qy        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
Db        553 TTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGTT 612

Qy        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
Db        613 CAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTTC 672

Qy        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
Db        673 AGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCAG 732

Qy        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
Db        733 GACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGTG 792

Qy        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
Db        793 ATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGGG 852

Qy        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903
Db        853 AAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCG 912

Qy        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963
Db        913 CAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTC 972

Qy        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023
Db        973 GCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGAG 1032

Qy       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083
Db       1033 GCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCG 1092

Qy       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143
Db       1093 CGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAG 1152

Qy       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203
Db       1153 GTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTAC 1212

Qy       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263
Db       1213 TACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTG 1272

Qy       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323
Db       1273 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGC 1332

Qy       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
Db       1333 TCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTC 1392

Qy       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443
Db       1393 TCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAAC 1452

Qy       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503
Db       1453 AACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTAC 1512

Qy       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563
Db       1513 GTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACACA 1572

Qy       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623
Db       1573 CTGGATGATGGTAGGGCCTCTGAACTCGCAGTGGGC------------------------ 1608

Qy       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683
Db       1609 ------------------------AGCTTGGGCCTGAGCAATGGGCAGGTGGTGAAGTAC 1644

Qy       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743
Db       1645 AAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAAT 1704

Qy       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803
Db       1705 GAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCG 1764

Qy       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863
Db       1765 GCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGT 1824

Qy       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923
Db       1825 ATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATG 1884

Qy       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983
Db       1885 GAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAAA 1944

Qy       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043
```

```
               || || ||| | ||||||||  | |||||||||||  |||| || ||||  ||||||||
Db        1945 CTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGGC 2004

Qy        2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103
               ||||||| || |||||||| || ||||| || ||||| ||||||||||||||||||||
Db        2005 CGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTAT 2064

Qy        2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163
               || ||||||||| |||||||||| ||| ||||||||||||||||| |||||||||||
Db        2065 GAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAA 2124

Qy        2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAG 2221
               || ||||| ||||| || || |||||||||||||||||||||||||||   | ||||
Db        2125 CATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAG 2182
```

RESULT 7
AAD10023
ID   AAD10023 standard; cDNA; 825 BP.
XX
AC   AAD10023;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) intracellular domain cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; intracellular domain; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..825
FT   /*tag= a
FT   /product= "Mouse Nope intracellular domain"
FT   /note= "CDS does not include start and stop codon"
FT   /partial
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAE05253.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 11; Page 83-85; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) intracellular domain cDNA.
XX
SQ   Sequence 825 BP; 174 A; 271 C; 236 G; 144 T; 0 other;

```
  Query Match             13.4%;  Score 825;  DB 22;  Length 825;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-204;
  Matches  825;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2932 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 2991
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 60

Qy        2992 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 3051
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 120

Qy        3052 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 3111
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 180

Qy        3112 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 3171
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 240

Qy        3172 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 3231
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 300

Qy        3232 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 3291
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 360

Qy        3292 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 3351
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 420

Qy        3352 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 3411
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 480

Qy        3412 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 3471
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 540

Qy        3472 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 3531
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 600

Qy        3532 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 3591
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 660

Qy        3592 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 3651
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 720

Qy        3652 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 3711
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 780

Qy        3712 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
               |||||||||||||||||||||||||||||||||||||||||||||
Db         781 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 825
```

```
RESULT 8
AAC79893
ID   AAC79893 standard; cDNA; 680 BP.
XX
AC   AAC79893;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Human secreted protein encoding cDNA for gene 45.
XX
KW   Human; secreted protein; cytostatic; antiarthritic; antiasthmatic;
KW   immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;
KW   neuroprotective; antidiabetic; tranquiliser; vulnerary; antibacterial;
KW   antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;
KW   autoimmune disorder; allergic condition; cardiovascular disorder;
KW   cancer; neurological disease; tissue repair; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055176-A2.
XX
PD   21-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06057.
XX
PR   12-MAR-1999;   99US-0124142.
PR   11-JUN-1999;   99US-0138597.
PR   03-DEC-1999;   99US-0168666.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-638176/61.
DR   P-PSDB; AAB44874.
XX
PT   Novel 49 human secreted proteins useful for diagnosis, prevention and
PT   treatment of disorders including neurological, cell proliferative,
PT   cardiovascular, and autoimmune/inflammatory disorders and microbial
PT   infections  -
XX
PS   Claim 1a; Page 358; 405pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) comprising an
CC   amino acid sequence at least 95 % identical to a polypeptide sequence
CC   selected from 49 polypeptides encoded by polynucleotide sequences
CC   included in American Type Culture Collection (ATCC) deposit number
CC   203917, defined in the specification. The products of the invention have
CC   cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic;
CC   antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,
CC   tranquiliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,
CC   antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)
CC   encoding (I) is useful for preventing, treating or ameliorating a medical
CC   condition and for diagnosing a pathological condition or susceptibility
CC   to the condition. (I) is useful for identifying a binding partner which
CC   affects the activity of the polypeptide and for identifying an activity
CC   in a biological sample. (I), (II) or an antibody (IV) specific to (I) is
CC   also useful for treating or preventing a disease, disorder or condition
CC   associated with aberrant expression of (I). Diseases treated or diagnosed
CC   include immune disorders such as autoimmune diseases, blood protein
CC   disorders, anemia, allergic reactions and conditions such as asthma,
CC   organ rejection or graft-versus-host disease, inflammation, hyper
CC   proliferative disorders, cardiovascular disorders such as arterioarterial
CC   fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ
CC   regeneration, cancer, neovascular glaucoma, diabetic retinopathy,
CC   rheumatoid arthritis, psoriasis, diseases associated with increased
CC   apoptosis that include acquired immunodeficiency syndrome (AIDS),
CC   neurological diseases such as Parkinson's disease, viral, bacterial,
CC   fungal or parasitic diseases.  They are also used to repair, replace or
CC   protect tissue damage by congenital defects, to treat trauma, in surgery,
CC   including cosmetic plastic surgery, to treat fibrosis, reperfusion injury
CC   or systemic cytokine damage, to stimulate chondrocyte growth, to prevent
CC   skin aging due to sunburn, to change a mammal's mental state or physical
CC   state by influencing biorhythms, cardiac rhythms, depression, memory,
CC   stress and for accelerating wound healing. (I), (II) and/or their agonist
CC   or antagonist are useful as food additives or preservatives to increase
CC   or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamin, mineral or other nutritional components. (I) is
CC   useful for screening therapeutic compounds. (II) is useful in forensic
CC   biology for detecting DNA sequences and as diagnostic probes for
CC   detecting the presence of specific mRNA in a particular cell type.
XX
SQ   Sequence 680 BP; 251 A; 97 C; 100 G; 232 T; 0 other;

  Query Match             5.5%;  Score 337.2;  DB 21;  Length 680;
  Best Local Similarity   79.9%;  Pred. No. 5.5e-77;
  Matches  473;  Conservative    0;  Mismatches  108;  Indels   11;  Gaps    6;

Qy       5588 GAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCATTCCTCAT 5647
              |||| ||||  | |||||||||||||||||||||||| | ||||||||||||||  ||||
Db         11 GAATTAAATGGCTGTTTTAACAGTAGGCTCTTAGCATTATACCACATAGTCATTTTTCAT 70

Qy       5648 GTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAATGAGACA 5707
              |||||||||||||| |||||    | |   ||||||||||||||||||||||||||||||
Db         71 GTTCTTGTTTAACAGGCACT--GAGGTTCTGGTTTAAATTAAATAGCTGCAAATGAGACA 128

Qy       5708 ATTTATAACCCATTAGGCTGGGTGGAAAATTG-TTCTCAAAAGCAAATAAGTAATAAATC 5766
              ||||||||||||||||| |||||||||||||| |||||||||||||||||||||||||||
Db        129 ATTTATAACCCATTAGGTTGGGTGGAAAATTGTTTCTCAAAAGCAAATAAGTAATAAATC 188

Qy       5767 TGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCATTATAT 5826
              |||||||||||||||||||||||||||||||||||| ||||   ||||||||||||||||
Db        189 TGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTGGCCATTTCTCACTAGCATTATAT 248

Qy       5827 ATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACCAGGTCT 5886
              |||||| ||| |||| |||| | || |||||||| |||||| ||||||||||  |  | |
Db        249 ATGATTTGGGCTCTGGGTAATTTGGAAGTGTTAGGTTTGTGTCTTTGTAGCAGTATTTTT 308

Qy       5887 TATTAGGAA-AGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCCATGGA 5945
                |    || | ||| |||||||||||||||| |||| ||||||||||   | |||| ||
Db        309 ATTAGAAAAGAATCTATTGGCCTTTTACAGGGTATTAATCCCTTTGTCACCTACCATTGA 368

Qy       5946 TGCCTTAAGTTCTTTGGAGTCTCATTTAAGAA-TTCCTTTTCTCGAAGCATGACAAGTGT 6004
              ||||||||||| | ||       ||| || |  |||||||||| || |||||||||||||
Db        369 TGCCTTAAGTTTTCTGAGTCTCAATTAAAAATCTTCCTTTTCTTGATGCATGACAAGTGT 428

Qy       6005 ATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTTAATT 6064
                   |||   ||| |||||| |  |||  |||  ||   ||      |     || | |
Db        429 -----AATCAGTACTTGCTCATTTATTTGTCTGTATTTAGTTTATGCTGTACTATTTAAT 483

Qy       6065 GCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACACTAAG 6124
                 | ||||||| |  |  || |||||||||||||||||| |||||| |||||  |||||
Db        484 TATCCTTCCAGCGTTTTTTTTCTCCTTACAAATATGATACTCTTTAGTGTTAAGCTAAG 543

Qy       6125 GTGTTGA-TCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGA 6175
              |  |||| |||||||||||||| | ||| |||||||||||||||||||||||
Db        544 GCATTGATTCATGTATCTGTCCTTATAATGAATTAATAAACTATTTTCCAGA 595


RESULT 9
ABX71437
ID   ABX71437 standard; cDNA; 2976 BP.
XX
AC   ABX71437;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP6.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
```

```
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54205.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 49-52; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 2976 BP; 844 A; 697 C; 653 G; 781 T; 1 other;

  Query Match             5.1%;  Score 316.6;  DB 25;  Length 2976;
  Best Local Similarity   47.4%;  Pred. No. 3.1e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db         45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db        579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db        639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db        699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db        759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db        819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db        879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db        933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db        993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352

Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412

Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472

Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
```

```
Db     1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532

Qy     1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db     1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592

Qy     1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db     1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652

Qy     1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db     1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712

Qy     1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db     1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1754

Qy     2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db     1755 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 1781

Qy     2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db     1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841

Qy     2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db     1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898

Qy     2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db     1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958

Qy     2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db     1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018

Qy     2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db     2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078

Qy     2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db     2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138

Qy     2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db     2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198

Qy     2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db     2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258

Qy     2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db     2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318

Qy     2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db     2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378

Qy     2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db     2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438

Qy     2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db     2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498

Qy     2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db     2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

```
RESULT 10
ABX71432
ID   ABX71432 standard; cDNA; 3210 BP.
XX
AC   ABX71432;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP1.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein;
KW   SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       76
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       706
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54200.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 13-18; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3210 BP; 895 A; 761 C; 716 G; 836 T; 2 other;

  Query Match             5.1%;  Score 316.6;  DB 25;  Length 3210;
```

```
Best Local Similarity   47.4%;  Pred. No. 3.2e-71;
Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db        279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db        813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db        873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db        933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db        993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db       1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db       1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db       1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db       1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586

Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646

Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706

Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db       1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766

Qy       1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db       1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826

Qy       1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db       1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886

Qy       1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db       1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946

Qy       1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db       1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1988

Qy       2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db       1989 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015

Qy       2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db       2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075

Qy       2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db       2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132

Qy       2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db       2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2192

Qy       2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db       2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252

Qy       2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db       2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312

Qy       2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db       2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372
```

```
Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db       2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432

Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db       2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492

Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db       2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552

Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db       2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612

Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db       2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672

Qy       2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db       2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732

Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db       2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765


RESULT 11
ABX71440
ID   ABX71440 standard; cDNA; 3219 BP.
XX
AC   ABX71440;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP9.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54208.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 65-68; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3219 BP; 927 A; 757 C; 706 G; 828 T; 1 other;

  Query Match             5.1%;  Score 316.6;  DB 25;  Length 3219;
  Best Local Similarity   47.4%;  Pred. No. 3.2e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db         45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db        579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db        639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db        699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db        759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818
```

```
Qy     1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db      819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy     1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db      879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy     1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db      933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy     1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db      993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy     1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db     1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy     1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db     1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy     1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db     1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy     1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db     1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy     1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db     1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352

Qy     1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db     1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412

Qy     1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db     1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472

Qy     1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db     1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532

Qy     1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db     1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592

Qy     1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db     1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652

Qy     1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db     1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712

Qy     1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db     1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1754

Qy     2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db     1755 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 1781

Qy     2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db     1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841

Qy     2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db     1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898

Qy     2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db     1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958

Qy     2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db     1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018

Qy     2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db     2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078

Qy     2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db     2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138

Qy     2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db     2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198

Qy     2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db     2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258

Qy     2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db     2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318

Qy     2634 GGAGTATCTAATTCTCTACAGCAACAACCCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db     2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378

Qy     2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db     2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438

Qy     2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db     2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498

Qy     2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db     2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

```
RESULT 12
ABX71447
ID   ABX71447 standard; cDNA; 3874 BP.
XX
AC   ABX71447;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
```

```
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 107-112; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3874 BP; 1077 A; 938 C; 897 G; 960 T; 2 other;

  Query Match             5.1%;  Score 316.6;  DB 25;  Length 3874;
  Best Local Similarity   47.4%;  Pred. No. 3.6e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db        469 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 528
Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        529 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 588
Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        589 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 648
Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        649 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 708
Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        709 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 768
Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        769 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 828
Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTGGAGGCTACCAG 716
Db        829 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 882
Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        883 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 942
Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        943 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 1002
Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db       1003 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 1062
Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db       1063 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTGTCGGGCCACTACCCCTGGCAC 1122
Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db       1123 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 1182
Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db       1183 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1242
Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db       1243 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAAATGGAAGGAAGATACATTCGAATGG 1302
Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db       1303 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1356
Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db       1357 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1416
Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db       1417 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1476
Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1477 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1536
Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1537 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1596
Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1597 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1656
Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1657 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1716
Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1717 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1776
Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1777 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1836
Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1837 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1896
Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db       1897 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1956
Qy       1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db       1957 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 2016
Qy       1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
```

```
Db       2017 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 2076

Qy       1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973

Db       2077 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 2136

Qy       1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033

Db       2137 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 2178

Qy       2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093

Db       2179 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2205

Qy       2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153

Db       2206 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2265

Qy       2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213

Db       2266 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2322

Qy       2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273

Db       2323 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2382

Qy       2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333

Db       2383 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2442

Qy       2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393

Db       2443 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2502

Qy       2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453

Db       2503 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2562

Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513

Db       2563 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2622

Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573

Db       2623 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2682

Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633

Db       2683 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2742

Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693

Db       2743 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2802

Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753

Db       2803 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2862

Qy       2754 TTTCTTCAAGATGGAGGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813

Db       2863 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2922

Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846

Db       2923 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2955
```

RESULT 13
AAD36777

```
ID   AAD36777 standard; cDNA; 2700 BP.
XX
AC   AAD36777;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human immunoglobulin superfamily protein (IGSFP)-1 cDNA.
XX
KW   Human; immunoglobulin superfamily protein-1; IGSFP-1; asthma;
KW   immune system disorder; acquired immune deficiency syndrome; AIDS;
KW   atherosclerosis; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; developmental disorder; renal tubular acidosis;
KW   anaemia; muscle disorder; cardiomyopathy; myocarditis; cancer;
KW   cell proliferative disorder; arteriosclerosis; hepatitis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2382
FT                   /*tag=  a
FT                   /product= "IGSFP-1"
FT   sig_peptide     1..105
FT                   /*tag=  b
FT   mat_peptide     106..2379
FT                   /*tag=  c
FT                   /product= "Mature IGSFP-1"
XX
PN   WO200240671-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44974.
XX
PR   16-NOV-2000; 2000US-249645P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Lu DAM,  Yue H,  Elliott VS,  Thangavelu K,  Ramkumar J;
PI   Lu Y,  Lo TP,  Gururajan R,  Gandhi AR,  Arvizu C,  Yao MG;
XX
DR   WPI; 2002-519384/55.
DR   P-PSDB; AAE14781.
XX
PT   Novel human immunoglobulin superfamily polypeptide, useful in
PT   diagnosis, prevention or treatment of immune system, neurological,
PT   developmental, muscle and cell proliferative disorders  -
XX
PS   Claim 5; Page 116-117; 122pp; English.
XX
CC   The present sequence is human immunoglobulin superfamily protein
CC   (IGSFP)-1 cDNA. The IGSFP polypeptide and polynucleotide are useful
CC   for diagnosing, treating or preventing disorders associated with aberrant
CC   expression of IGSFP e.g. immune system disorders (e.g. acquired immune
CC   deficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis, uveitis),
CC   neurological disorders (e.g. Alzheimer's disease, Huntington's disease,
CC   dementia, Parkinson's disease), developmental disorders (e.g. renal
CC   tubular acidosis, epilepsy, anaemia), muscle disorders (e.g.
CC   cardiomyopathy, myocarditis), or cell proliferative disorders (e.g.
CC   arteriosclerosis, cirrhosis, hepatitis, and cancer). The polypeptide and
CC   polynucleotide are also useful for assessing the effects of exogenous
CC   compounds on their expression. The polypeptide is useful in drug
CC   screening techniques, to analyse the proteome of a tissue or cell type,
CC   as elements on a microarray. The polynucleotide is useful for creating
CC   knock-in humanised animals or transgenic animals to model human diseases,
CC   in somatic or germline gene therapy, to generate a transcript image of a
CC   tissue or cell type, for detecting differences in the chromosomal
CC   location due to translocation, inversion among normal, carrier or
CC   affected individuals, and as hybridisation probes for mapping naturally
CC   occurring genomic sequences.
XX
SQ   Sequence 2700 BP; 528 A; 859 C; 815 G; 498 T; 0 other;

  Query Match             5.1%;  Score 316.4;  DB 24;  Length 2700;
  Best Local Similarity   55.0%;  Pred. No. 3.3e-71;
  Matches  692;  Conservative    0;  Mismatches  551;  Indels   15;  Gaps    3;

Qy        342 TGAGGGCAGCTATTCCTGTCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGC 401
```

```
Db     333 TGAAGGTGACTATGAGTGTGTGGCCCAGAACCGCTTTGGGCTGGTGGTCAGCCGGAAGGC 392

Qy     402 TGTGGTCAAGCTTGCCACACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGA 461
Db     393 TCGCATCCAAGCTGCAACCATGTCGGACTTCCACGTGCATCCCCAGGCCACCGTGGGTGA 452

Qy     462 GGAGAACGGGACAGCACGCTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTAC 521
Db     453 GGAGGGTGGTGTGGCCCGCTTCCAGTGCCAAATCCATGGGCTTCCCAAACCCCTGATCAC 512

Qy     522 TTGGGAAAAGGACCAGGT---GACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAA 578
Db     513 TTGGGAGAAGAACAGAGTCCCAATTGACACGGACAATGAGAGGTACACATTGCTGCCCAA 572

Qy     579 GTGGCTCCTCCAGATCCTAGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGC 638
Db     573 GGGGGTCCTGCAGATCACAGGACTTCGAGCTGAGGACGGTGGCATCTTCCACTGTGTGGC 632

Qy     639 CACCAATTCAGCCCGCCAACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGG 698
Db     633 CTCAAACATCGCCAGTATCCGGATCAGCCACGGGGCCAGGCTCACTGTGTCAGGCTCGGG 692

Qy     699 GTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCAC 758
Db     693 CTCTGGGGCCTACA---------AGGAGCCAGCCATCCTCGTGGGGCCTGAGAACCTCAC 743

Qy     759 GGTAGTGTCTGGACAGAATGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTT 818
Db     744 CCTGACAGTGCACCAGACCGCGGTGCTTGAGTGTGTCGCCACGGGCAACCCGCGCCCCAT 803

Qy     819 TGTGTCCTGGGTCCGACAGGATGGAAAGCCTATCTCCACGGATGTCATC---GTTCTGGG 875
Db     804 TGTGTCCTGGAGCCGCCTGGATGGTCGCCCTATCGGGGTGGAGGGCATCCAGGTGCTGGG 863

Qy     876 CCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCG 935
Db     864 CACAGGAAACCTCATCATCTCAGACGTGACGGTCCAGCACTCTGGCGTCTACGTCTGTGC 923

Qy     936 AGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGC 995
Db     924 AGCCAACAGACCTGGCACCCGGGTGAGGAGAACGGCACAGGGCCGGCTGGTGGTGCAAGC 983

Qy     996 TGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCG 1055
Db     984 CCCAGCTGAGTTTGTGCAGCATCCCCAGTCCATCTCCAGGCCAGCTGGGACCACAGCCAT 1043

Qy    1056 CTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGAT 1115
Db    1044 GTTCACCTGCCAAGCCCAGGGTGAGCCACCGCCTCATGTCACGTGGCTGAAAAATGGACA 1103

Qy    1116 CCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCAC 1175
Db    1104 GGTGCTGGGGCCAGGAGGCCACGTCAGGCTCAAGAATAACAACAGCACACTGACCATTTC 1163

Qy    1176 TCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAAC 1235
Db    1164 TGGAATCGGTCCTGAGGATGAAGCCATTTATCAGTGTGTGGCCGAGAACAGTGCGGGCTC 1223

Qy    1236 TGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGAC 1295
Db    1224 ATCACAGGCCAGTGCCAGGCTGACCGTACTGTGGGCTGAGGGGCTCCCCGGGCCTCCCCG 1283

Qy    1296 TCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGA 1355
Db    1284 CAATGTGCGGGCAGTCTCTGTGTCTTCCACTGAGGTGCGTGTGTCCTGGAGTGAGCCGCT 1343

Qy    1356 GTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGA 1415
Db    1344 GGCCAACACCAAGGAGATCATCGGCTACGTCCTGCACATCAGGAAGGCTGCTGACCCACC 1403

Qy    1416 CAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCT 1475
Db    1404 GGAGCTGGAGTATCAGGAGGCAGTCAGCAAGAGCACCTTTCAGCACCTGGTCAGCGACCT 1463

Qy    1476 GGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCG 1535
Db    1464 GGAGCCCTCCACAGCCTACAGTTTCTACATCAAGGCCTACACACCAAGGGGGGCCAGCTC 1523

Qy    1536 AACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAG 1593
Db    1524 AGCCTCTGTGCCCACCCTAGCTAGCACCCTGGGTGAAGCCCCTGCCCCACCCCCACTG 1581
```

RESULT 14
ABX71435
ID   ABX71435 standard; cDNA; 3453 BP.
XX
AC   ABX71435;
XX
DT   06-MAR-2003 (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP4.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW, Scoville J, Turner CA, Friedrich G, Abuin A;
PI   Zambrowicz B, Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54203.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays -
XX
PS   Claim 1; Column 31-36; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3453 BP; 979 A; 821 C; 769 G; 883 T; 1 other;

Query Match 5.1%; Score 315.4; DB 25; Length 3453;

```
Best Local Similarity   47.4%;  Pred. No. 6.9e-71;
Matches 1210;  Conservative    0;  Mismatches 1271;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db        279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        693 CTTCCACACACCAACAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db        813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db        873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db        933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db        993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db       1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db       1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db       1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db       1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586

Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646

Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706

Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db       1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766

Qy       1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db       1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826

Qy       1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db       1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886

Qy       1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db       1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946

Qy       1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db       1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA----------------- 1988

Qy       2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db       1989 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015

Qy       2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db       2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075

Qy       2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db       2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132

Qy       2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db       2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCACCATCTCTATGCGAA 2192

Qy       2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db       2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252

Qy       2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db       2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312

Qy       2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db       2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372
```

```
Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db       2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432

Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db       2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492

Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db       2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552

Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db       2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612

Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db       2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672

Qy       2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGGA 2813
Db       2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732

Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db       2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765
```

RESULT 15
AAS76339
ID   AAS76339 standard; cDNA; 2695 BP.
XX
AC   AAS76339;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12143.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12152.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12143; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2695 BP; 542 A; 822 C; 852 G; 479 T; 0 other;

Query Match 4.7%; Score 289; DB 23; Length 2695;
Best Local Similarity 54.8%; Pred. No. 4.4e-64;
Matches 643; Conservative 0; Mismatches 515; Indels 15; Gaps 3;

```
Qy        427 GACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTTGAA 486
Db        211 GACTTCCACGTGCATCCCCAGGCCACCGTGGGTGAGGAGGGTGGTGTGGCCCGCTTCCAG 270

Qy        487 TGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACCGTG 546
Db        271 TGCCAAATCCATGGGCTTCCCAAACCCCTGATCACTTGGGAGAAGAACAGAGTCCCAATT 330

Qy        547 C---CTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
Db        331 GACACGGACAATGAGAGGTACACATTGCTGCCCAAGGGGGTCCTGCAGATCACAGGACTT 390

Qy        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
Db        391 CGAGCTGAGGACGGTGGCATCTTCCACTGTGTGGCCTCAAACATCGCCAGTATCCGGATC 450

Qy        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
Db        451 AGCCACGGGGCCAGGCTCACTGTGTCAGGCTCGGGCTCTGGGGCCTACA---------AG 501

Qy        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
Db        502 GAGCCAGCCATCCTCGTGGGGCCTGAGAACCTCACCCTGACAGTGCACCAGACCGCGGTG 561

Qy        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
Db        562 CTTGAGTGTGTCGCCACGGGCAACCCGCGCCCCATTGTGTCCTGGAGCCGCCTGGATGGT 621

Qy        844 AAGCCTATCTCCACGGATGTCATC---GTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
Db        622 CGCCCTATCGGGGTGGAGGGCATCCAGGTGCTGGGCACAGGAAACCTCATCATCTCAGAC 681

Qy        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
Db        682 GTGACGGTCCAGCACTCTGGCGTCTACGTCTGTGCAGCCAACAGACCTGGCACCCGGGTG 741

Qy        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
Db        742 AGGAGAACGGCACAGGGCCGGCTGGTGGTGCAAGCCCCAGCTGAGTTTGTGCAGCATCCC 801

Qy       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
Db        802 CAGTCCATCTCCAGGCCAGCTGGGACCACAGCCATGTTCACCTGCCAAGCCCAGGGTGAG 861

Qy       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
Db        862 CCACCGCCTCATGTCACGTGGCTGAAAAATGGACAGGTGCTGGGGCCAGGAGGCCACGTC 921

Qy       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
```

```
Db        922 AGGCTCAAGAATAACAACAGCACACTGACCATTTCTGGAATCGGTCCTGAGGATGAAGCC 981

Qy       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Db        982 ATTTATCAGTGTGTGGCCGAGAACAGTGCGGGCTCATCACAGGCCAGTGCCAGGCTGACC 1041

Qy       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Db       1042 GTACTGTGGGCTGAGGGGCTCCCCGGGCCTCCCCGCAATGTGCGGGCAGTCTCTGTGTCT 1101

Qy       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Db       1102 TCCACTGAGGTGCGTGTGTCCTGGAGTGAGCCGCTGGCCAACACCAAGGAGATCATCGGC 1161

Qy       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Db       1162 TACGTCCTGCACATCAGGAAGGCTGCTGACCCACCGGAGCTGGAGTATCAGGAGGCAGTC 1221

Qy       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Db       1222 AGCAAGAGCACCTTTCAGCACCTGGTCAGCGACCTGGAGCCCTCCACAGCCTACAGTTTC 1281

Qy       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Db       1282 TACATCAAGGCCTACACACCAAGGGGGGCCAGCTCAGCCTCTGTGCCCACCCTAGCTAGC 1341

Qy       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAG 1593

Db       1342 ACCCTGGGTGAAGCCCCTGCCCCACCCCCACTG 1374
```

Search completed: October 14, 2003, 10:17:18
Job time : 1181.58 secs

```
OM nucleic - nucleic search, using sw model

Run on:          October 14, 2003, 17:10:36 ; Search time 1668.34 Seconds
                                          (without alignments)
                                          9605.664 Million cell updates/sec

Title:           US-09-754-997A-1
Perfect score:   6176
Sequence:        1 atggcgcgggcggacacggg..........taataaactattttccagac 6176

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters:      3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                 11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                 12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                 14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                 15:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 16:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                 17:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6176	100.0	6176	10	US-09-754-997A-1	Sequence 1, Appli
2	3756	60.8	3756	10	US-09-754-997A-45	Sequence 45, Appl
3	2804.2	45.4	3753	12	US-10-239-663-29	Sequence 29, Appl
4	2796	45.3	2796	10	US-09-754-997A-3	Sequence 3, Appli
5	2681.8	43.4	3741	10	US-09-908-193-1	Sequence 1, Appli
6	1522.4	24.7	2223	12	US-10-239-663-28	Sequence 28, Appl
7	825	13.4	825	10	US-09-754-997A-5	Sequence 5, Appli
8	316.6	5.1	2976	14	US-10-231-353-11	Sequence 11, Appl
9	316.6	5.1	3210	14	US-10-231-353-1	Sequence 1, Appli
10	316.6	5.1	3219	14	US-10-231-353-17	Sequence 17, Appl
11	316.6	5.1	3874	14	US-10-231-353-31	Sequence 31, Appl
12	315.4	5.1	3453	14	US-10-231-353-7	Sequence 7, Appli
13	288	4.7	288	10	US-09-754-997A-19	Sequence 19, Appl
14	252	4.1	252	10	US-09-754-997A-23	Sequence 23, Appl
15	249	4.0	249	10	US-09-754-997A-15	Sequence 15, Appl
16	249	4.0	249	10	US-09-754-997A-17	Sequence 17, Appl
17	246	4.0	246	10	US-09-754-997A-21	Sequence 21, Appl
18	243	3.9	243	10	US-09-754-997A-7	Sequence 7, Appli
19	217	3.5	2481	14	US-10-231-353-15	Sequence 15, Appl
20	217	3.5	2715	14	US-10-231-353-5	Sequence 5, Appli
21	217	3.5	2724	14	US-10-231-353-19	Sequence 19, Appl
22	217	3.5	2958	14	US-10-231-353-9	Sequence 9, Appli
23	213.8	3.5	2139	14	US-10-231-353-21	Sequence 21, Appl
24	213.8	3.5	2382	14	US-10-231-353-27	Sequence 27, Appl
25	195	3.2	195	10	US-09-754-997A-13	Sequence 13, Appl
26	192	3.1	192	10	US-09-754-997A-9	Sequence 9, Appli
27	189	3.1	189	10	US-09-754-997A-11	Sequence 11, Appl
28	151.2	2.4	1875	14	US-10-231-353-23	Sequence 23, Appl
29	122.2	2.0	249	10	US-09-933-797-25	Sequence 25, Appl
30	119	1.9	909	14	US-10-231-353-13	Sequence 13, Appl
31	119	1.9	1143	14	US-10-231-353-3	Sequence 3, Appli
32	117	1.9	1644	14	US-10-231-353-25	Sequence 25, Appl
33	117	1.9	1887	14	US-10-231-353-29	Sequence 29, Appl
34	85.6	1.4	439	10	US-09-728-445-856	Sequence 856, App
35	71	1.1	4277	11	US-09-946-374-57	Sequence 57, Appl
36	71	1.1	4277	12	US-10-015-387A-57	Sequence 57, Appl
37	71	1.1	4277	12	US-10-006-130A-57	Sequence 57, Appl
38	71	1.1	4277	12	US-10-199-672-439	Sequence 439, App
39	71	1.1	4277	12	US-10-006-172A-57	Sequence 57, Appl
40	71	1.1	4277	12	US-10-187-749-439	Sequence 439, App
41	71	1.1	4277	12	US-10-194-457-439	Sequence 439, App
42	71	1.1	4277	12	US-10-184-642-439	Sequence 439, App
43	71	1.1	4277	12	US-10-196-747-439	Sequence 439, App
44	71	1.1	4277	12	US-10-015-392A-57	Sequence 57, Appl
45	71	1.1	4277	12	US-10-017-253A-57	Sequence 57, Appl

ALIGNMENTS

```
RESULT 1
US-09-754-997A-1
; Sequence 1, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 6176
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(3756)
US-09-754-997A-1

  Query Match             100.0%;  Score 6176;  DB 10;  Length 6176;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6176;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy       181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy       241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy       301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy       361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy       421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy       481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy       541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy       601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy       661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy       721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy       781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy       841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy       901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy       961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy      1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy      1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy      1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy      1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy      1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy      1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy      1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy      1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy      1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy      1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy      1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy      1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy      1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy      1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy      1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy      1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy      1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy      2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy      2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy      2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy      2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy      2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
```

```
Qy  2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy  2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy  2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy  2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy  2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy  2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy  2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy  2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy  2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy  2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy  2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy  3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy  3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy  3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy  3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy  3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy  3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy  3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy  3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy  3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy  3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy  3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy  3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy  3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCCTGAGCAGAAGGCAGATATGGCTCA 3780

Qy  3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781 GGAACATGCCATGCATGGCTACACATGTGTGTACTAGAGATATCCATAAGTCCTTGGAGC 3840

Qy  3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841 CTCTTAGGGTCCTTTGGCTGGGGTTGGGGAGAACTTTACTCTCCCTCATATTCTGCATCA 3900

Qy  3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901 CATACAGGAGGGACTTGAGACACAGCTCTGTGTAATGGACACGTGTGAAGTCGTGTGTGT 3960

Qy  3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961 GTGTGTGTGTGTGTGTGCTGGTTGAGCTAGGAAACCTCTCCCTATGTAGCACTCACTGTG 4020

Qy  4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021 GCCTAGTTGACCCTCCGTGGCAGGATGGTGTAACAGTGATCAGTGCCAGCTCTTTGAGCT 4080

Qy  4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081 TTTAGCCTTGTCACCTAGCCTTTTATTACACTCTGAGAGTGTCTCCAGTGCTGTGTCTAC 4140

Qy  4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141 AAAGACAGCGCCCAGCCCTCTTCTGTCAGCTGTGCTGAGCAGAGTGCCAGTCAACTCCAC 4200

Qy  4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201 GGGCCTATGACACCGCAGCCTACCACAGCATGGCTGTCATCCCCCTGGCCTCCTAAGGTC 4260

Qy  4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261 CAGATGTCTGGGTGAACCCAGCTCAGCTCCCCTCTCCTTTGAGCATCTCTGTACCTAATT 4320

Qy  4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4321 TTGTAATCTGGGAAGTGCCTGGTTTGGGAAATCTTCTTTCGCACCCTGTCCCTCTCTGCC 4380

Qy  4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381 CCTTCCTTCATTTGTTCTGGTGATCTGTCTCATGTCATCTTGCTCGATTATCCTGGGGCC 4440

Qy  4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441 CTTCTCTTTCCCATGATGCCCCTGATTTCCTCACTGCTGTTTTCATTTCTGTCTGCCATG 4500

Qy  4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 CTTGTCTTTATGTCGTGTGTTTCTCGTCCCTGAGTTCAACCTATGCACCCTTTCCTAACA 4560

Qy       4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 ACATGACTACCTCATGTCTGCTTCAGACCATAGTGTGACCCCTGGGTCCCCACAGCTCCC 4620

Qy       4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CTGCCAACCGCCTTCCTGGGCAGATGAGCCCACTCCAAGTAGATCTGGAAAAGACCCTTG 4680

Qy       4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 TGGCTTGTCTGGCTGCCCTCCCCTTGGTGTTGAGATGAGAAGGTTTTCTATGGAAGAGAT 4740

Qy       4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GAGTCCAGGCTGCACAGGGGAACCCCCAAGAAGGGGTAGGGAGTGAAACCAAGAGGCTGA 4800

Qy       4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 AAAAAAATGGCTGCCACCCATCTGCACAGAGAGATGGGTGTGTGCTTTTGACGTGCAGTC 4860

Qy       4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 CTGGCTGAAACTGAAGGGGTGAGGAGAGGGGAGCTACTGGGGCTGCCATGGCTCAGTTCC 4920

Qy       4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 CTGACCCTGGAGCCCTGAACCTGGCTTCAGAGTAGCAAAGAGTTTCCTCCAAGATGCTGT 4980

Qy       4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4981 AAGGGAAGTCTTTGCATAGGAAAAGGGCGGCTGGCTCATTTTATTTTATCTTTCTTTACA 5040

Qy       5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 CTGAATCCCAAAATCATCTTACCACAAAGGGCCAAGCCTGACTGGTATTTCCTGAGTCAC 5100

Qy       5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 AAGAGCCATGCCATCTCTCTGGTTTCTCACCTCAGTCATGTCCCAGAATTGTCAGATCCA 5160

Qy       5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5161 GTGGCATCTGTGCTCTTGCTGCACATCTTTCTATTTCAACTGGCTGGCACATCAAGTGTT 5220

Qy       5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5221 AACTCTGGCTTCTGGGCCAAGTTAGAAATAACCAGTCTATTTTCCCTTTATTTTATTTTA 5280

Qy       5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5281 TTTTATTTTATTTTATGTCTTTCAGTGGAGTTGTAGCTTCTGAAAGCGTCTGTGTTTATT 5340

Qy       5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 AGCCTTGTGTGTCACTCATGTTTGACCCCACCCACATTTCCTTCTCCTCCCCTCTTCAGC 5400

Qy       5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 CAGCCTATGATAACACTAAAGATTATTAATGCTGGCTTCGTATCTCATTAAAGACAGGAT 5460

Qy       5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 TGTCACTTGAACTACTTCTATAGCATTCAAAGTGGCCACGGCCAACACCACCGTATGTTT 5520

Qy       5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 CTTCATTGCTCTGAAGGTCAAGAGCCTCATTTTGTTTTCCTGGTTAGATTCTTTTCCTCC 5580

Qy       5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 TTGCCTTGAATGAAATAACCGTTTTAACAGTAGGCTCTTAGCATCACACCACATAGTCAT 5640

Qy       5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 TCCTCATGTTCTTGTTTAACAAGCACTTGAGGTTCTGGGTTTAAATTAAATAGCTGCAAA 5700

Qy       5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 TGAGACAATTTATAACCCATTAGGCTGGGTGGAAAATTGTTCTCAAAAGCAAATAAGTAA 5760

Qy       5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 TAAATCTGGTATCTGCCTATAACTCACAGTTGATAAGAAAGTAGCCAGAACTCACTAGCA 5820

Qy       5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5821 TTATATATGATTGGGGTTCTGAGTAACTGGGGAGTGTTAGCTTTGTGACTTTGTAGCACC 5880

Qy       5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5881 AGGTCTTATTAGGAAAGTCTGTTGGCCTTTTACAGGGCATTAGTCCCTTTGTCGTTTGCC 5940

Qy       5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5941 ATGGATGCCTTAAGTTCTTTGGAGTCTCATTTAAGAATTCCTTTTCTCGAAGCATGACAA 6000

Qy       6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6001 GTGTATCGCAATACTTACATGCTCACTCGTTTACCTGGCTTAGTTTGTGCTGGGTTATTT 6060

Qy       6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6061 AATTGCACTTTCCAGCATCATGCTTCCTCCTTACAAATATGATATTCTTTATTGTTACAC 6120

Qy       6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6121 TAAGGTGTTGATCATGTATCTGTCCCTGTAAAGAATTAATAAACTATTTTCCAGAC 6176

RESULT 2
US-09-754-997A-45
; Sequence 45, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 3756
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-754-997A-45

  Query Match             60.8%;  Score 3756;  DB 10;  Length 3756;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3756;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120
```

```
Qy    121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy    181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy    241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy    301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy    361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy    421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy    481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy    541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy    601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy    661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy    721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy    781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy    841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy    901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy    961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy   1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy   1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy   1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200
```

```
Qy   1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy   1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy   1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy   1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy   1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy   1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy   1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy   1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy   1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy   1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy   1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy   1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy   2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy   2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy   2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy   2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy   2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
              ||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 3
US-10-239-663-29
; Sequence 29, Application US/10239663
; Publication No. US20030139572A1
; GENERAL INFORMATION:
;  APPLICANT: Agarwal, Pankaj
;  APPLICANT:  Murdoch, Paul R.
;  APPLICANT:  Rizvi, Safia, K.
;  APPLICANT:  Smith, Randall, F.
;  APPLICANT:  Xiang, Zhaoying
;  APPLICANT:  Kabnick, Karen
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP50018
;  CURRENT APPLICATION NUMBER: US/10/239,663
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/US01/09226
;  PRIOR FILING DATE: 2001-03-22
;  PRIOR APPLICATION NUMBER: 60/192,158
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/192,668
;  PRIOR FILING DATE: 2000-03-27
;  PRIOR APPLICATION NUMBER: 60/200,166
;  PRIOR FILING DATE: 2000-04-27
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 3753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-239-663-29

  Query Match             45.4%;  Score 2804.2;  DB 12;  Length 3753;
  Best Local Similarity   84.7%;  Pred. No. 0;
  Matches 3171;  Conservative    0;  Mismatches  568;  Indels    6;  Gaps    2;

Qy          3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62
              || |    || | |    ||||||||||||||| | | ||||||||||||| || |||||||
Db          9 GGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 68

Qy         63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122
              |||||||||||  ||||||||||||||| ||||| |||||||||||||    | ||| || ||
Db         69 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACT 128

Qy        123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182
              ||||||||||||||||||| ||||||||||  ||||| |||| |   ||||||||| | ||
Db        129 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 188
```

```
Qy    183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242
Db    189 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 248

Qy    243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302
Db    249 CCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 308

Qy    303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362
Db    309 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGTCATTGAAGGCAACTATTCGTGCCT 365

Qy    363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422
Db    366 AGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACT 425

Qy    423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482
Db    426 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 485

Qy    483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542
Db    486 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 545

Qy    543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602
Db    546 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 605

Qy    603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662
Db    606 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 665

Qy    663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722
Db    666 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCA 725

Qy    723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782
Db    726 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 785

Qy    783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842
Db    786 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGG 845

Qy    843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902
Db    846 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 905

Qy    903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962
Db    906 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 965

Qy    963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGA 1022
Db    966 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGA 1025

Qy   1023 GGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCC 1082
Db   1026 GGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCC 1085

Qy   1083 ACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAA 1142
Db   1086 GCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAA 1145

Qy   1143 GGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTA 1202
Db   1146 GGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTA 1205

Qy   1203 CTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGT 1262
Db   1206 CTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGT 1265

Qy   1263 AGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAG 1322
Db   1266 GGTGGTGCGCGAGGGGCTGCCCAGCGCCCCACGCGGGTCACTGCTACGCCACTGAGCAG 1325

Qy   1323 CTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTT 1382
Db   1326 CTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTT 1385

Qy   1383 CTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAA 1442
Db   1386 CTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAA 1445

Qy   1443 CAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTA 1502
Db   1446 CAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTA 1505

Qy   1503 CGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATAC 1562
Db   1506 CGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACAC 1565

Qy   1563 ACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGA 1622
Db   1566 ACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGA 1625

Qy   1623 CATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTA 1682
Db   1626 CATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTA 1685

Qy   1683 CAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAA 1742
Db   1686 CAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAA 1745

Qy   1743 TGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTC 1802
Db   1746 TGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTC 1805

Qy   1803 AGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGG 1862
Db   1806 GGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAG 1865

Qy   1863 TGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGAT 1922
Db   1866 TATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGAT 1925

Qy   1923 GGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAA 1982
Db   1926 GGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAA 1985

Qy   1983 ACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGG 2042
Db   1986 ACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGG 2045

Qy   2043 TCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTA 2102
Db   2046 CCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTA 2105

Qy   2103 TGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAA 2162
Db   2106 TGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAA 2165

Qy   2163 ACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGA 2222
Db   2166 ACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGA 2225

Qy   2223 CCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAG 2282
Db   2226 CATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAG 2285

Qy   2283 CTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAA 2342
Db   2286 CTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAA 2345

Qy   2343 CTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATAC 2402
```

```
Db       2346 CTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACAC 2405

Qy       2403 CAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGC 2462
Db       2406 CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGC 2465

Qy       2463 GGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCAC 2522
Db       2466 AGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCAC 2525

Qy       2523 CCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATC 2582
Db       2526 CCTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTC 2585

Qy       2583 CACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCT 2642
Db       2586 CACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCT 2645

Qy       2643 AATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGA 2702
Db       2646 GATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCA 2705

Qy       2703 GGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAA 2762
Db       2706 GGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAA 2765

Qy       2763 GATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTAC 2822
Db       2766 GATGGGGGCGCGCACAGAGGTGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCAC 2825

Qy       2823 TCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTGGG 2882
Db       2826 GCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATCATCGTGGG 2885

Qy       2883 TGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGCTC 2942
Db       2886 TGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCC 2945

Qy       2943 CCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTCTA 3002
Db       2946 CCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTA 3005

Qy       3003 CACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGT 3062
Db       3006 CTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGT 3065

Qy       3063 GCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGT 3122
Db       3066 GCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGT 3125

Qy       3123 ACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTC 3182
Db       3126 GCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGATCTC 3185

Qy       3183 CTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCA 3242
Db       3186 CTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCA 3245

Qy       3243 GGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCA 3302
Db       3246 GG---CAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCA 3302

Qy       3303 GACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAGCC 3362
Db       3303 GACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGGAGGAAGAAGTC 3362

Qy       3363 GTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGC 3422
Db       3363 ACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGC 3422

Qy       3423 ATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGC 3482
Db       3423 ATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCC 3482

Qy       3483 AGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGG 3542
Db       3483 AGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGA 3542

Qy       3543 CAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCC 3602
Db       3543 CAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCC 3602

Qy       3603 AGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGA 3662
Db       3603 AGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGA 3662

Qy       3663 GGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGCCT 3722
Db       3663 GACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCT 3722

Qy       3723 TCCCAGGGCCCCTGTCTCCTCTGCT 3747
Db       3723 GCCCAGATCCCCGGTCTCCTCCTCT 3747
```

RESULT 4
US-09-754-997A-3
; Sequence 3, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
; APPLICANT: Salbaum, Michael J.
; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
; TITLE OF INVENTION: Acids and Methods of Use
; FILE REFERENCE: P-NI 4552
; CURRENT APPLICATION NUMBER: US/09/754,997A
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,496
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: US 60/205,789
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 2796
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2796)
US-09-754-997A-3

```
  Query Match             45.3%;  Score 2796;  DB 10;  Length 2796;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123
Db          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183
Db         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243
Db        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303
Db        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363
Db        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
```

```
Qy        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGACAGCACGCTTT 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGACAGCACGCTTT 420

Qy        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy       2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy       2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy       2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy       2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy       2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy       2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy       2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643

Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy       2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703

Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy       2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763

Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy       2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823

Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy       2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859

Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

```
RESULT 5
US-09-908-193-1
; Sequence 1, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
;  APPLICANT: RASTELLI, LUCA
;  APPLICANT:  SHIMKETS, RICHARD A.
;  APPLICANT:  ZERHUSEN, BRYAN
;  APPLICANT:  MALYANKAR, URIEL M.
;  APPLICANT:  PADIGARU, MURALIDHARA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
;  FILE REFERENCE: 21402-062
;  CURRENT APPLICATION NUMBER: US/09/908,193
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: 60/220,273
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: 60/221,650
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: 60/221,233
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: 60/220,912
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/218,875
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/218,870
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/218,901
;  PRIOR FILING DATE: 2000-07-18
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 3741
;   TYPE: DNA
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
;   OTHER INFORMATION: with homology to NOPE-like protein
;   NAME/KEY: modified_base
;   LOCATION: (375)
;   OTHER INFORMATION: a, c, t, g, other or unknown
US-09-908-193-1

  Query Match             43.4%;  Score 2681.8;  DB 10;  Length 3741;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 3141;  Conservative    0;  Mismatches  583;  Indels   30;  Gaps    8;

Qy          3 GGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCGCG 62

Db          9 GGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTGGCCGCGCG 68

Qy         63 CGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCT 122

Db         69 CGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGCCACT 128

Qy        123 GCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGC 182

Db        129 GCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGC 188

Qy        183 TGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGA 242

Db        189 CGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGA 248

Qy        243 GAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGA 302

Db        249 CCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGG 308

Qy        303 CAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCT 362

Db        309 CAG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCT 365

Qy        363 GGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACT 422

Db        366 AGCCCACGGNCCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCAGTCT 425

Qy        423 CGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTT 482

Db        426 CGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTT 485

Qy        483 TGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGAC 542

Db        486 TGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGAC 545

Qy        543 CGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGT 602

Db        546 ATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGT 605

Qy        603 CCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATT 662

Db        606 TCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTT 665

Qy        663 CAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCA 722

Db        666 CAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACCAGGGGGCA 725

Qy        723 GGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGT 782

Db        726 GGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGT 785

Qy        783 GATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGG 842

Db        786 GATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA---GACGG 842

Qy        843 AAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGC 902

Db        843 GAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGC 902

Qy        903 GCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTT 962

Db        903 GCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTT 962

Qy        963 CGCCACTGCGGCTGCTGAGCTCCGAGTGCTT---GCTGCCCCAGCCATCTCGCAGGCACC 1019

Db        963 CGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACTCAGGCGCC 1022

Qy       1020 CGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGA 1079

Db       1023 CGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGA 1082

Qy       1080 GCCACGGCCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGT 1139

Db       1083 GCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGT 1142

Qy       1140 CAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGG 1199
```

```
Db   1143 CAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGG 1202

Qy   1200 CTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGC 1259
Db   1203 CTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGC 1262

Qy   1260 GGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAG 1319
Db   1263 CGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAG 1322

Qy   1320 CAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGG 1379
Db   1323 CAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGG 1382

Qy   1380 CTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGT 1439
Db   1383 CTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAGTTTGCAGT 1442

Qy   1440 AAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTT 1499
Db   1443 GAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTT 1502

Qy   1500 CTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCA 1559
Db   1503 CTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCA 1562

Qy   1560 TACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTC 1619
Db   1563 CACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTC 1622

Qy   1620 GGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAA 1679
Db   1623 GGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAA 1682

Qy   1680 GTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGG 1739
Db   1683 GTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGG 1742

Qy   1740 AAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGAT 1799
Db   1743 AAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGAT 1802

Qy   1800 TTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACC 1859
Db   1803 TTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCC 1862

Qy   1860 TGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAA 1919
Db   1863 CAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAA 1922

Qy   1920 GATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATA 1979
Db   1923 GATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTA 1982

Qy   1980 CAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGG 2039
Db   1983 CAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGG 2042

Qy   2040 GGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCA 2099
Db   2043 GGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCA 2102

Qy   2100 GTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAA 2159
Db   2103 GTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAA 2162

Qy   2160 CAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCC 2219
Db   2163 CAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACC 2222

Qy   2220 AGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAA 2279
Db   2223 AGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAA 2282

Qy   2280 CAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGT 2339
Db   2283 CAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGT 2342

Qy   2340 CAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTA 2399
Db   2343 CAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTA 2402

Qy   2400 TACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTT 2459
Db   2403 ---CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTT 2459

Qy   2460 TGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTC 2519
Db   2460 TGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTC 2519

Qy   2520 CACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACC 2579
Db   2520 CACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACC 2579

Qy   2580 ATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTA 2639
Db   2580 GTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTA 2639

Qy   2640 TCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCAC 2699
Db   2640 TCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCAC 2699

Qy   2700 ------AGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2700 GCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTA 2759

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2760 CTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGA 2819

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCAT 2873
Db   2820 TGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCAT 2879

Qy   2874 CATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACG 2933
Db   2880 CATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCG 2939

Qy   2934 ACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCC 2993
Db   2940 CCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCC 2999

Qy   2994 AGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGA 3053
Db   3000 CGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCAGCCCCCCAGCTGCCCATGAATTGGA 3059

Qy   3054 GTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAA 3113
Db   3060 GTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAG 3119

Qy   3114 GGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAG 3173
Db   3120 GGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAG 3179

Qy   3174 AAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGA 3233
Db   3180 AAAGGT---AAGTGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGA 3236

Qy   3234 GCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGG 3293
Db   3237 GCTGCCCCAGG---CAGGCCCCCGGCCGGCTCTGACCCGGCCCTGCTGCCCCCTGCTGG 3293

Qy   3294 AACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAG 3353
Db   3294 AACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGATC------AGGGCAATGGGAG 3347
```

```
Qy     3354 AAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGA 3413
Db     3348 GAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGA 3407

Qy     3414 CTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACC 3473
Db     3408 CTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCC 3467

Qy     3474 ACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGA 3533
Db     3468 CCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGC 3527

Qy     3534 CTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCAC 3593
Db     3528 CTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTAC 3587

Qy     3594 CTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGC 3653
Db     3588 CTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGT 3647

Qy     3654 TATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAG 3713
Db     3648 GCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAG 3707

Qy     3714 CCCAAGCCTTCCCAGGGCCCTGTCTCCTCTGCT 3747
Db     3708 CCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741
```

RESULT 6
US-10-239-663-28
; Sequence 28, Application US/10239663
; Publication No. US20030139572A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kabnick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50018
; CURRENT APPLICATION NUMBER: US/10/239,663
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-239-663-28

```
  Query Match             24.7%;  Score 1522.4;  DB 12;  Length 2223;
  Best Local Similarity   82.8%;  Pred. No. 0;
  Matches 1786;  Conservative    0;  Mismatches  321;  Indels   51;  Gaps    2;

Qy       64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123
Db       76 GGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACTG 135

Qy      124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183
Db      136 CAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGCC 195

Qy      184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243
Db      196 GCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGAC 255

Qy      244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303
Db      256 CACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGGC 315

Qy      304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363
Db      316 AG---TGACGAGTCAGTCCCTGAGGCTGTGGGGTCATTGAAGGCAACTATTCGTGCCTA 372

Qy      364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423
Db      373 GCCCACGGCCCCCCTGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACTC 432

Qy      424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483
Db      433 GCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTTT 492

Qy      484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543
Db      493 GAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGACA 552

Qy      544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603
Db      553 TTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGTT 612

Qy      604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
Db      613 CAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTTC 672

Qy      664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723
Db      673 AGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCAG 732

Qy      724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783
Db      733 GACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGTG 792

Qy      784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843
Db      793 ATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGGG 852

Qy      844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903
Db      853 AAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCG 912

Qy      904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963
Db      913 CAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTC 972

Qy      964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023
Db      973 GCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGAG 1032

Qy     1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083
Db     1033 GCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCG 1092

Qy     1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143
Db     1093 CGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAG 1152

Qy     1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203
Db     1153 GTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTAC 1212

Qy     1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263
Db     1213 TACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTG 1272

Qy     1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323
```

```
Db       1273 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGC 1332

Qy       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
Db       1333 TCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTC 1392

Qy       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443
Db       1393 TCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAAC 1452

Qy       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503
Db       1453 AACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTAC 1512

Qy       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563
Db       1513 GTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACACA 1572

Qy       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623
Db       1573 CTGGATGATGGTAGGGCCTCTGAACTCGCAGTGGGC------------------------ 1608

Qy       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683
Db       1609 ------------------------AGCTTGGGCCTGAGCAATGGGCAGGTGGTGAAGTAC 1644

Qy       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743
Db       1645 AAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAAT 1704

Qy       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803
Db       1705 GAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCG 1764

Qy       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863
Db       1765 GCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGT 1824

Qy       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923
Db       1825 ATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATG 1884

Qy       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983
Db       1885 GAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAAA 1944

Qy       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043
Db       1945 CTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGGC 2004

Qy       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103
Db       2005 CGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTAT 2064

Qy       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163
Db       2065 GAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAA 2124

Qy       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAG 2221
Db       2125 CATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAG 2182
```

RESULT 7
US-09-754-997A-5
; Sequence 5, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
; APPLICANT: Salbaum, Michael J.
; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
; TITLE OF INVENTION: Acids and Methods of Use
; FILE REFERENCE: P-NI 4552
; CURRENT APPLICATION NUMBER: US/09/754,997A
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,496
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: US 60/205,789
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(825)
US-09-754-997A-5

```
  Query Match             13.4%;  Score 825;  DB 10;  Length 825;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-237;
  Matches  825;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2932 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 2991
Db          1 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 60

Qy       2992 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 3051
Db         61 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 120

Qy       3052 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 3111
Db        121 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 180

Qy       3112 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 3171
Db        181 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 240

Qy       3172 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 3231
Db        241 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 300

Qy       3232 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 3291
Db        301 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 360

Qy       3292 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 3351
Db        361 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 420

Qy       3352 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 3411
Db        421 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 480

Qy       3412 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 3471
Db        481 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 540

Qy       3472 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 3531
Db        541 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 600

Qy       3532 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 3591
Db        601 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 660

Qy       3592 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 3651
Db        661 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 720

Qy       3652 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 3711
Db        721 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 780
```

```
Qy       3712 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
Db        781 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 825

RESULT 8
US-10-231-353-11
; Sequence 11, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Enco
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/10/231,353
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/877,730
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 2976
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-11

  Query Match              5.1%;  Score 316.6;  DB 14;  Length 2976;
  Best Local Similarity   47.4%;  Pred. No. 1.3e-83;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db         45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db        579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db        639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db        699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db        759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db        819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db        879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db        933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db        993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352

Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412

Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472

Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db       1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532

Qy       1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db       1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592

Qy       1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db       1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652

Qy       1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db       1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712
```

```
Qy       1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033

Db       1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1754

Qy       2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093

Db       1755 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 1781

Qy       2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153

Db       1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841

Qy       2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213

Db       1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898

Qy       2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273

Db       1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958

Qy       2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333

Db       1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018

Qy       2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393

Db       2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078

Qy       2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453

Db       2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138

Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513

Db       2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198

Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573

Db       2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258

Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633

Db       2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318

Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693

Db       2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378

Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753

Db       2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438

Qy       2754 TTTCTTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813

Db       2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498

Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846

Db       2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

RESULT 9
US-10-231-353-1
; Sequence 1, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064Alel Human Phosphatases and Polynucleotides En
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3210
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-1

Query Match 5.1%; Score 316.6; DB 14; Length 3210;
Best Local Similarity 47.4%; Pred. No. 1.4e-83;
Matches 1210; Conservative 1; Mismatches 1270; Indels 72; Gaps 6;

```
Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db        279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db        339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db        399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db        459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db        519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db        579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db        639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db        693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db        753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db        813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db        873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db        933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db        993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
```

```
Db   1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db   1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db   1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db   1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db   1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db   1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db   1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db   1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy   1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db   1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1988

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   1989 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2192

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db   2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372

Qy   2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db   2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432

Qy   2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db   2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492

Qy   2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db   2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552

Qy   2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db   2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612

Qy   2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db   2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765
```

RESULT 10
US-10-231-353-17
; Sequence 17, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides E
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 3219
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-17

Query Match 5.1%; Score 316.6; DB 14; Length 3219;

```
Best Local Similarity   47.4%;  Pred. No. 1.4e-83;
Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy    300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db     45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy    360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db    105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy    420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db    165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy    480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db    225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy    540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db    285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db    345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db    405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db    459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db    519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db    579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db    639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db    699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db    759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db    819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db    879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db    933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db    993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db   1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db   1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db   1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db   1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy   1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db   1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1754

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   1755 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 1781

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db   2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138
```

```
Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db       2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198

Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db       2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258

Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db       2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318

Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db       2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378

Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db       2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438

Qy       2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db       2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498

Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db       2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

RESULT 11
US-10-231-353-31
; Sequence 31, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Encc
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 3874
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-31

```
  Query Match             5.1%;  Score 316.6;  DB 14;  Length 3874;
  Best Local Similarity   47.4%;  Pred. No. 1.6e-83;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db        469 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 528

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        529 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 588

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        589 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 648

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        649 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGAGTTCAATCGGAC 708

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        709 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 768

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        769 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 828

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        829 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 882

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db        883 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 942

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db        943 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 1002

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db       1003 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 1062

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db       1063 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 1122

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db       1123 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 1182

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db       1183 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1242

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db       1243 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1302

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db       1303 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1356

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db       1357 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1416

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db       1417 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1476

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db       1477 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1536

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db       1537 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1596

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db       1597 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1656

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db       1657 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1716

Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
```

```
Db   1717 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1776

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1777 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1836

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1837 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1896

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1897 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1956

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1957 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 2016

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   2017 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 2076

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   2077 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 2136

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   2137 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 2178

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   2179 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2205

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   2206 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2265

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   2266 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2322

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   2323 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2382

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   2383 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2442

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2443 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2502

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db   2503 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2562

Qy   2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db   2563 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2622

Qy   2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db   2623 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2682

Qy   2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db   2683 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2742

Qy   2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db   2743 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2802

Qy   2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2803 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2862

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2863 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2922

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db   2923 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2955
```

RESULT 12
US-10-231-353-7
; Sequence 7, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides En
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3453
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-231-353-7

Query Match 5.1%; Score 315.4; DB 14; Length 3453;
Best Local Similarity 47.4%; Pred. No. 3.4e-83;
Matches 1210; Conservative 0; Mismatches 1271; Indels 72; Gaps 6;

```
Qy    300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db    279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy    360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db    339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy    420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db    399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy    480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db    459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy    540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db    519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db    579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db    639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
```

```
Db    693 CTTCCACACACCAACAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db    753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db    813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db    873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db    933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db    993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db   1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db   1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db   1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db   1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db   1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db   1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db   1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db   1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy   1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db   1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1988

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   1989 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2192

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db   2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372

Qy   2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db   2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432

Qy   2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db   2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492

Qy   2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db   2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552

Qy   2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db   2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612

Qy   2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db   2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db   2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765
```

RESULT 13
US-09-754-997A-19

```
; Sequence 19, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 288
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(288)
US-09-754-997A-19

  Query Match             4.7%;  Score 288;  DB 10;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 1e-75;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1888 TTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAG 1947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAG 60

Qy       1948 CCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACA 2007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACA 120

Qy       2008 GAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTC 2067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTC 180

Qy       2068 GGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGC 2127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGC 240

Qy       2128 AGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGC 2175
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGC 288

RESULT 14
US-09-754-997A-23
; Sequence 23, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 252
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(252)
US-09-754-997A-23

  Query Match             4.1%;  Score 252;  DB 10;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-65;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2542 ACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGG 2601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGG 60

Qy       2602 TGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAAC 2661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAAC 120

Qy       2662 CACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCA 2721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCA 180

Qy       2722 GAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAG 2781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAG 240

Qy       2782 GTGGGGCCTGGG 2793
              ||||||||||||
Db        241 GTGGGGCCTGGG 252

RESULT 15
US-09-754-997A-15
; Sequence 15, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 249
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(249)
US-09-754-997A-15

  Query Match             4.0%;  Score 249;  DB 10;  Length 249;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-64;
  Matches  249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1285 AGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGG 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGG 60

Qy       1345 GAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCA 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCA 120

Qy       1405 AGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAG 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAG 180

Qy       1465 GTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTG 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTG 240
```

```
Qy       1525 GGGGCCAGC 1533
              |||||||||
Db        241 GGGGCCAGC 249

Search completed: October 15, 2003, 12:28:06
Job time : 1679.34 secs
```

OM nucleic - nucleic search, using sw model

Run on: October 14, 2003, 17:02:36 ; Search time 371.754 Seconds
(without alignments)
7332.764 Million cell updates/sec

Title: US-09-754-997A-1
Perfect score: 6176
Sequence: 1 atggcgcgggcggacacggg..........taataaactattttccagac 6176

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	316.6	5.1	2976	4	US-09-877-730-11	Sequence 11, Appl
	2	316.6	5.1	3210	4	US-09-877-730-1	Sequence 1, Appli
	3	316.6	5.1	3219	4	US-09-877-730-17	Sequence 17, Appl
	4	316.6	5.1	3874	4	US-09-877-730-31	Sequence 31, Appl
	5	315.4	5.1	3453	4	US-09-877-730-7	Sequence 7, Appli
	6	217	3.5	2481	4	US-09-877-730-15	Sequence 15, Appl
	7	217	3.5	2715	4	US-09-877-730-5	Sequence 5, Appli
	8	217	3.5	2724	4	US-09-877-730-19	Sequence 19, Appl
	9	217	3.5	2958	4	US-09-877-730-9	Sequence 9, Appli
	10	213.8	3.5	2139	4	US-09-877-730-21	Sequence 21, Appl
	11	213.8	3.5	2382	4	US-09-877-730-27	Sequence 27, Appl
	12	151.2	2.4	1875	4	US-09-877-730-23	Sequence 23, Appl
	13	119	1.9	909	4	US-09-877-730-13	Sequence 13, Appl
	14	119	1.9	1143	4	US-09-877-730-3	Sequence 3, Appli
	15	117	1.9	1644	4	US-09-877-730-25	Sequence 25, Appl
	16	117	1.9	1887	4	US-09-877-730-29	Sequence 29, Appl
	17	86.8	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	18	48	0.8	2049	4	US-09-099-749-10	Sequence 10, Appl
	19	46	0.7	2600	2	US-08-427-497E-4	Sequence 4, Appli
	20	46	0.7	3189	2	US-08-427-497E-3	Sequence 3, Appli
	21	46	0.7	3774	2	US-08-341-843B-1	Sequence 1, Appli
	22	46	0.7	3774	2	US-08-427-497E-1	Sequence 1, Appli
	23	46	0.7	3774	2	US-08-427-497E-2	Sequence 2, Appli
	24	46	0.7	3888	4	US-08-506-296B-13	Sequence 13, Appl
c	25	45.6	0.7	1447	4	US-09-443-041A-27	Sequence 27, Appl
c	26	45.4	0.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	27	44.8	0.7	5993	3	US-09-383-630-1	Sequence 1, Appli
	28	44.8	0.7	5993	3	US-09-383-630-2	Sequence 2, Appli
	29	44.4	0.7	1004	3	US-09-383-586-24	Sequence 24, Appl
	30	44.4	0.7	3402	4	US-09-996-243-118	Sequence 118, App
	31	43.2	0.7	4078	4	US-09-016-434-1132	Sequence 1132, Ap
	32	43.2	0.7	6000	1	US-08-348-006B-6	Sequence 6, Appli
	33	43.2	0.7	6000	2	US-08-800-825A-6	Sequence 6, Appli
	34	43.2	0.7	6000	3	US-09-158-657-6	Sequence 6, Appli
	35	43.2	0.7	6000	5	PCT-US94-10166-6	Sequence 6, Appli
	36	43	0.7	1134	4	US-09-252-991A-4496	Sequence 4496, Ap
	37	43	0.7	1764	4	US-09-252-991A-4599	Sequence 4599, Ap
c	38	43	0.7	1863	4	US-09-252-991A-4262	Sequence 4262, Ap
	39	42.8	0.7	6814	4	US-09-484-970B-66	Sequence 66, Appl
	40	42.6	0.7	1042	2	US-08-427-497E-6	Sequence 6, Appli
	41	42	0.7	636	4	US-09-252-991A-1835	Sequence 1835, Ap
	42	42	0.7	925	4	US-09-257-179-17	Sequence 17, Appl
c	43	42	0.7	1272	4	US-09-252-991A-2123	Sequence 2123, Ap
c	44	41.8	0.7	6961	1	US-08-727-034-2	Sequence 2, Appli
	45	41.8	0.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-877-730-11
; Sequence 11, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2976
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-11

```
  Query Match             5.1%;  Score 316.6;  DB 4;  Length 2976;
  Best Local Similarity   47.4%;  Pred. No. 1.1e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
              | ||| |  ||| | |||||      |      | ||  |  ||| ||    |||   ||
Db         45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
                ||||    | |      ||||   |    || ||    |||    |     |  | ||
Db        105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
                |    |  ||     | || ||    |||      ||  | ||    ||    || ||
Db        165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGAAAAGGACCAGGT 539
               |||| |||| | |       ||  ||  |  |  |||| || ||||     | | |
Db        225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGAGTTCAATCGGAC 284

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
               ||  | |||  |     |  | || | |   | || || |   |  |  | ||||||
Db        285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344
```

```
Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db    345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db    405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db    459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db    519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db    579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db    639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db    699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db    759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db    819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db    879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db    933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db    993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db   1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db   1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db   1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db   1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy   1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db   1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA----------------- 1754

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   1755 ---------------------------------TGGCCCATTTTCTTGGATACCAAGGA 1781

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db   2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138

Qy   2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db   2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198

Qy   2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db   2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258

Qy   2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db   2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318

Qy   2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db   2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378

Qy   2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db   2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438

Qy   2754 TTTCTTCAAGATGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
```

```
Db     2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498

Qy     2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846

Db     2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

RESULT 2
US-09-877-730-1
; Sequence 1, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3210
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-1

```
  Query Match             5.1%;  Score 316.6;  DB 4;  Length 3210;
  Best Local Similarity   47.4%;  Pred. No. 1.1e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy      300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db      279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy      360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db      339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy      420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db      399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy      480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db      459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy      540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db      519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy      597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db      579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy      657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db      639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy      717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db      693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy      777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db      753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy      837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db      813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy      894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db      873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy      954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db      933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy     1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db      993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy     1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db     1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy     1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193

Db     1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy     1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253

Db     1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy     1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313

Db     1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy     1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373

Db     1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy     1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433

Db     1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy     1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493

Db     1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy     1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553

Db     1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy     1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613

Db     1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586

Qy     1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673

Db     1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646

Qy     1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733

Db     1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706

Qy     1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793

Db     1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766

Qy     1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853

Db     1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826

Qy     1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913

Db     1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886

Qy     1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
```

```
Db     1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946

Qy     1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033

Db     1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1988

Qy     2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093

Db     1989 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015

Qy     2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153

Db     2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075

Qy     2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213

Db     2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132

Qy     2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273

Db     2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2192

Qy     2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333

Db     2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252

Qy     2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393

Db     2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312

Qy     2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453

Db     2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372

Qy     2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513

Db     2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432

Qy     2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573

Db     2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492

Qy     2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633

Db     2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552

Qy     2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693

Db     2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612

Qy     2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753

Db     2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672

Qy     2754 TTTCTTCAAGATGGAGGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813

Db     2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732

Qy     2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846

Db     2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765
```

RESULT 3
US-09-877-730-17
; Sequence 17, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 3219
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-17

```
  Query Match             5.1%;  Score 316.6;  DB 4;  Length 3219;
  Best Local Similarity   47.4%;  Pred. No. 1.1e-71;
  Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy      300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db       45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy      360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db      105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy      420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db      165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy      480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db      225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy      540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db      285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy      597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db      345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy      657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db      405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 458

Qy      717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db      459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy      777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db      519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy      837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db      579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy      894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db      639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy      954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db      699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy     1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db      759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818
```

```
Qy     1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db      819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878
Qy     1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db      879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932
Qy     1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db      933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992
Qy     1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db      993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052
Qy     1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db     1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112
Qy     1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db     1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172
Qy     1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db     1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232
Qy     1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db     1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292
Qy     1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db     1293 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1352
Qy     1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db     1353 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1412
Qy     1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db     1413 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1472
Qy     1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db     1473 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1532
Qy     1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db     1533 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1592
Qy     1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db     1593 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1652
Qy     1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db     1653 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1712
Qy     1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db     1713 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1754
Qy     2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db     1755 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 1781
Qy     2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db     1782 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 1841
Qy     2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db     1842 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 1898
Qy     2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db     1899 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 1958
Qy     2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db     1959 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2018
Qy     2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db     2019 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2078
Qy     2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db     2079 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2138
Qy     2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db     2139 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2198
Qy     2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db     2199 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2258
Qy     2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db     2259 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2318
Qy     2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db     2319 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2378
Qy     2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db     2379 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2438
Qy     2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGCCCCTTTTCCCGCTTGCAGGA 2813
Db     2439 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2498
Qy     2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db     2499 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2531
```

RESULT 4
US-09-877-730-31
; Sequence 31, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 3874
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-31

```
Query Match             5.1%;  Score 316.6;  DB 4;  Length 3874;
Best Local Similarity   47.4%;  Pred. No. 1.2e-71;
Matches 1210;  Conservative    1;  Mismatches 1270;  Indels   72;  Gaps    6;

Qy    300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db    469 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 528

Qy    360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db    529 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 588

Qy    420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db    589 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 648

Qy    480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db    649 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 708

Qy    540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db    709 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 768

Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db    769 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 828

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db    829 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 882

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db    883 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 942

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db    943 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 1002

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db   1003 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 1062

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db   1063 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTGTCGGGCCACTACCCCTGGCAC 1122

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db   1123 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 1182

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db   1183 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1242

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db   1243 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1302

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db   1303 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1356

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db   1357 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1416

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db   1417 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1476

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db   1477 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1536

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db   1537 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1596

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db   1597 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1656

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db   1657 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1716

Qy   1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db   1717 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1776

Qy   1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db   1777 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1836

Qy   1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db   1837 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1896

Qy   1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db   1897 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1956

Qy   1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db   1957 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 2016

Qy   1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db   2017 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 2076

Qy   1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db   2077 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 2136

Qy   1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db   2137 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 2178

Qy   2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db   2179 ---------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2205

Qy   2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db   2206 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2265

Qy   2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db   2266 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2322

Qy   2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db   2323 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2382

Qy   2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db   2383 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2442

Qy   2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db   2443 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2502

Qy   2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
```

```
Db   2503 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2562

Qy   2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513

Db   2563 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2622

Qy   2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573

Db   2623 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2682

Qy   2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633

Db   2683 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2742

Qy   2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693

Db   2743 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2802

Qy   2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753

Db   2803 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2862

Qy   2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813

Db   2863 CATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGA 2922

Qy   2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846

Db   2923 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2955
```

RESULT 5
US-09-877-730-7
; Sequence 7, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3453
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-7

```
  Query Match             5.1%;  Score 315.4;  DB 4;  Length 3453;
  Best Local Similarity   47.4%;  Pred. No. 2.3e-71;
  Matches 1210;  Conservative    0;  Mismatches 1271;  Indels   72;  Gaps    6;

Qy    300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db    279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTATCAGTG 338

Qy    360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db    339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy    420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db    399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy    480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db    459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy    540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db    519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db    579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db    639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTGATTCCAG------CTAAGGAGTCAAAATC 692

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db    693 CTTCCACACACCAACAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db    753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db    813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db    873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db    933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db    993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db   1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193

Db   1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253

Db   1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313

Db   1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373

Db   1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433

Db   1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493

Db   1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553

Db   1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526
```

```
Qy       1554 GGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAA 1613
Db       1527 ACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAG 1586
Qy       1614 CCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGT 1673
Db       1587 TCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGT 1646
Qy       1674 GCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGT 1733
Db       1647 GGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCT 1706
Qy       1734 GCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGT 1793
Db       1707 CCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGT 1766
Qy       1794 CCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAG 1853
Db       1767 TCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAG 1826
Qy       1854 GACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAG 1913
Db       1827 GACGCCCAAAGCTACAAGCGTGAAAGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCT 1886
Qy       1914 GGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTC 1973
Db       1887 GAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCA 1946
Qy       1974 TGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCC 2033
Db       1947 GGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAA------------------ 1988
Qy       2034 CCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGT 2093
Db       1989 ----------------------------------TGGGCCCATTTTCTTGGATACCAAGGA 2015
Qy       2094 GAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGC 2153
Db       2016 CCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGC 2075
Qy       2154 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 2213
Db       2076 TTACAACAACATAGACGATGGCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATG 2132
Qy       2214 CACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGA 2273
Db       2133 CGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAA 2192
Qy       2274 GTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAA 2333
Db       2193 GGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACA 2252
Qy       2334 GATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCAC 2393
Db       2253 AATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCT 2312
Qy       2394 CTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTA 2453
Db       2313 GTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATA 2372
Qy       2454 CGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGA 2513
Db       2373 CGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTA 2432
Qy       2514 ACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCT 2573
Db       2433 CCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAAT 2492
Qy       2574 GACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGT 2633
Db       2493 AGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGAC 2552
Qy       2634 GGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCT 2693
Db       2553 CCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTT 2612
Qy       2694 CACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTA 2753
Db       2613 ACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTA 2672
Qy       2754 TTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGA 2813
Db       2673 CATTGTCAAGATATCTGCATCCAATGAGGTGGAGAAGGACCCTTTTCAAATTCTGTGGA 2732
Qy       2814 TGTGATTACTCTGCAAGAGACATTCTCAGACTC 2846
Db       2733 GCTGGCAGTACTTCCAAAGGAAACCTCTGAATC 2765
```

RESULT 6
US-09-877-730-15
; Sequence 15, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 2481
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-15

```
  Query Match             3.5%;  Score 217;  DB 4;  Length 2481;
  Best Local Similarity   50.3%;  Pred. No. 5.8e-46;
  Matches  647;  Conservative    1;  Mismatches  621;  Indels   18;  Gaps    4;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db         45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTATCAGTG 104
Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db        105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164
Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db        165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224
Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db        225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284
Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db        285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344
Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db        345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404
Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db        405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 458
```

```
Qy     717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db     459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy     777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db     519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy     837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db     579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy     894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db     639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy     954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db     699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy    1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db     759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy    1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
Db     819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy    1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193
Db     879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy    1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253
Db     933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy    1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313
Db     993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy    1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373
Db    1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy    1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433
Db    1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy    1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493
Db    1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy    1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
Db    1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy    1554 GGTGCATACACTGGACGATGTCCCCAG 1580
Db    1293 ACAGAATACTCTAGAGGATGACCCCAG 1319
```

RESULT 7
US-09-877-730-5
; Sequence 5, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 2715
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-5

```
  Query Match             3.5%;  Score 217;  DB 4;  Length 2715;
  Best Local Similarity   50.3%;  Pred. No. 6.1e-46;
  Matches  647;  Conservative    1;  Mismatches  621;  Indels   18;  Gaps    4;

Qy     300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
Db     279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy     360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419
Db     339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy     420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479
Db     399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy     480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539
Db     459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy     540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596
Db     519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy     597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656
Db     579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy     657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716
Db     639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 692

Qy     717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776
Db     693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy     777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836
Db     753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy     837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893
Db     813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy     894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953
Db     873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy     954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013
Db     933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy    1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073
Db     993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy    1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133
```

```
Db   1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193

Db   1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253

Db   1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313

Db   1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373

Db   1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433

Db   1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493

Db   1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553

Db   1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy   1554 GGTGCATACACTGGACGATGTCCCCAG 1580

Db   1527 ACAGAATACTCTAGAGGATGACCCCAG 1553
```

RESULT 8
US-09-877-730-19
; Sequence 19, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 2724
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-19

```
  Query Match             3.5%;  Score 217;  DB 4;  Length 2724;
  Best Local Similarity   50.3%;  Pred. No. 6.1e-46;
  Matches  647;  Conservative    1;  Mismatches  621;  Indels   18;  Gaps    4;

Qy    300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db     45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy    360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db    105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy    420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db    165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy    480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db    225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy    540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db    285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy    597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db    345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy    657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db    405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 458

Qy    717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db    459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy    777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db    519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy    837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db    579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy    894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db    639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy    954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db    699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy   1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db    759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy   1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db    819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy   1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193

Db    879 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 932

Qy   1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253

Db    933 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 992

Qy   1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313

Db    993 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1052

Qy   1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373

Db   1053 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1112

Qy   1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433

Db   1113 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1172

Qy   1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493

Db   1173 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1232

Qy   1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553
```

```
Db       1233 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1292

Qy       1554 GGTGCATACACTGGACGATGTCCCCAG 1580

Db       1293 ACAGAATACTCTAGAGGATGACCCCAG 1319
```

RESULT 9
US-09-877-730-9
; Sequence 9, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2958
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-9

```
  Query Match             3.5%;  Score 217;  DB 4;  Length 2958;
  Best Local Similarity   50.3%;  Pred. No. 6.4e-46;
  Matches  647;  Conservative   1;  Mismatches  621;  Indels   18;  Gaps    4;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db        279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy        360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db        339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy        420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db        399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy        480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db        459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy        540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db        519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy        597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db        579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy        657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db        639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 692

Qy        717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db        693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy        777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db        753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy        837 GGATGGAAAGCCTAT---CTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db        813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy        894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db        873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy        954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db        933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy       1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db        993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy       1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db       1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy       1134 GCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGA 1193

Db       1113 TAGAATTAAAATGTACAACAGT------AAATTGGTAATTAACCAGATTATTCCTGAAGA 1166

Qy       1194 CGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCC 1253

Db       1167 TGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAG 1226

Qy       1254 CCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCC 1313

Db       1227 ACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAAC 1286

Qy       1314 GCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAAT 1373

Db       1287 CATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGT 1346

Qy       1374 CATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTT 1433

Db       1347 CATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGT 1406

Qy       1434 TGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTA 1493

Db       1407 AGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTA 1466

Qy       1494 TGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCT 1553

Db       1467 TACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGAC 1526

Qy       1554 GGTGCATACACTGGACGATGTCCCCAG 1580

Db       1527 ACAGAATACTCTAGAGGATGACCCCAG 1553
```

RESULT 10
US-09-877-730-21
; Sequence 21, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding :
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 2139
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-21

  Query Match             3.5%;  Score 213.8;  DB 4;  Length 2139;
  Best Local Similarity   47.0%;  Pred. No. 3.6e-45;
  Matches  793;  Conservative    0;  Mismatches  842;  Indels   54;  Gaps    2;

Qy       1158 CGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGC 1217
Db         60 CAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATGGC 119

Qy       1218 AGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGG 1277
Db        120 TGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAAGA 179

Qy       1278 GCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGT 1337
Db        180 CAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTTTT 239

Qy       1338 GGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCA 1397
Db        240 AGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTACAT 299

Qy       1398 AAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGA 1457
Db        300 GAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACTCA 359

Qy       1458 GCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTC 1517
Db        360 TTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATATAT 419

Qy       1518 CCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCC 1577
Db        420 GCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTTCC 479

Qy       1578 CAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATG 1637
Db        480 CCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCCTG 539

Qy       1638 GCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGG 1697
Db        540 GCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTCCG 599

Qy       1698 TTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTAC 1757
Db        600 CCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTACCT 659

Qy       1758 GTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGC 1817
Db        660 TTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGAGT 719

Qy       1818 TGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAG 1877
Db        720 GGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCTACAAGCGTGAA 779

Qy       1878 CCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGT 1937
Db        780 AGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCTGAACTGTACCACCATTTCTGTGAG 839

Qy       1938 GTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGA 1997
Db        840 GTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTACTACAAGGA 899

Qy       1998 GGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGC 2057
Db        900 AGAAGGGCAGCAGGAGAA------------------------------------------ 917

Qy       2058 TTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTT 2117
Db        918 ---------TGGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTCAGTGGCTT 968

Qy       2118 AGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTA 2177
Db        969 AGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGCTA 1028

Qy       2178 CGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAG 2237
Db       1029 TCAGGCAGATCAGACTGTCAGCAC---TCCAGGATGCGTGTCTGTTCGTGATCGCATGGT 1085

Qy       2238 GGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTG 2297
Db       1086 CCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTT 1145

Qy       2298 GCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTT 2357
Db       1146 CCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTG 1205

Qy       2358 CGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGA 2417
Db       1206 TAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCA 1265

Qy       2418 CATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGG 2477
Db       1266 CATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGT 1325

Qy       2478 AGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCC 2537
Db       1326 GGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACC 1385

Qy       2538 TTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACA 2597
Db       1386 AGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTC 1445

Qy       2598 CTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAA 2657
Db       1446 TTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATC 1505

Qy       2658 CAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAG 2717
Db       1506 TAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCAT 1565

Qy       2718 TGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCAC 2777
Db       1566 GGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAA 1625

Qy       2778 AGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATT 2837
Db       1626 TGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAAC 1685

Qy       2838 CTCAGACTC 2846
Db       1686 CTCTGAATC 1694

RESULT 11
US-09-877-730-27
; Sequence 27, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
```

```
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 2382
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-27

  Query Match              3.5%;  Score 213.8;  DB 4;  Length 2382;
  Best Local Similarity   47.0%;  Pred. No. 3.8e-45;
  Matches  793;  Conservative    0;  Mismatches  842;  Indels   54;  Gaps    2;

Qy       1158 CGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGC 1217
Db         60 CAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATGGC 119

Qy       1218 AGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGG 1277
Db        120 TGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAAGA 179

Qy       1278 GCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGT 1337
Db        180 CAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTTTT 239

Qy       1338 GGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCA 1397
Db        240 AGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTACAT 299

Qy       1398 AAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGA 1457
Db        300 GAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACTCA 359

Qy       1458 GCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTC 1517
Db        360 TTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATATAT 419

Qy       1518 CCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCC 1577
Db        420 GCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTTCC 479

Qy       1578 CAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATG 1637
Db        480 CCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCCTG 539

Qy       1638 GCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGG 1697
Db        540 GCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTCCG 599

Qy       1698 TTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTAC 1757
Db        600 CCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTACCT 659

Qy       1758 GTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGC 1817
Db        660 TTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGAGT 719

Qy       1818 TGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAG 1877
Db        720 GGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCTACAAGCGTGAA 779

Qy       1878 CCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGT 1937
Db        780 AGCCCCTAAGTCTCCAGAGTTGCATTTGGAGCCTCTGAACTGTACCACCATTTCTGTGAG 839

Qy       1938 GTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGA 1997
Db        840 GTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTACTACAAGGA 899

Qy       1998 GGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGC 2057
Db        900 AGAAGGGCAGCAGGAGAA------------------------------------------ 917

Qy       2058 TTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTT 2117
Db        918 ---------TGGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTCAGTGGCTT 968

Qy       2118 AGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTA 2177
Db        969 AGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGCTA 1028

Qy       2178 CGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAG 2237
Db       1029 TCAGGCAGATCAGACTGTCAGCAC---TCCAGGATGCGTGTCTGTTCGTGATCGCATGGT 1085

Qy       2238 GGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTG 2297
Db       1086 CCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTT 1145

Qy       2298 GCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTT 2357
Db       1146 CCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTG 1205

Qy       2358 CGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGA 2417
Db       1206 TAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCA 1265

Qy       2418 CATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGG 2477
Db       1266 CATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGT 1325

Qy       2478 AGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCC 2537
Db       1326 GGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACC 1385

Qy       2538 TTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACA 2597
Db       1386 AGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTC 1445

Qy       2598 CTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAA 2657
Db       1446 TTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATC 1505

Qy       2658 CAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAG 2717
Db       1506 TAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCAT 1565

Qy       2718 TGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCAC 2777
Db       1566 GGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAA 1625

Qy       2778 AGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATT 2837
Db       1626 TGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAAC 1685

Qy       2838 CTCAGACTC 2846
Db       1686 CTCTGAATC 1694
```

```
RESULT 12
US-09-877-730-23
; Sequence 23, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding =
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
```

```
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 1875
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-23

  Query Match             2.4%;  Score 151.2;  DB 4;  Length 1875;
  Best Local Similarity   45.8%;  Pred. No. 5.4e-29;
  Matches  662;  Conservative    0;  Mismatches  728;  Indels   54;  Gaps    2;

Qy       1403 CAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGC 1462
Db         41 CAACAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACTCATTATA 100

Qy       1463 AGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGC 1522
Db        101 TTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATATATGCCAA 160

Qy       1523 TGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCG 1582
Db        161 TGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTTCCCCTGA 220

Qy       1583 CAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGC 1642
Db        221 GACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCCTGGCTGC 280

Qy       1643 CCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGG 1702
Db        281 CAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAA 340

Qy       1703 GGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAA 1762
Db        341 GTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTACCTTTTGG 400

Qy       1763 ACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCT 1822
Db        401 AAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGAGTGGGGC 460

Qy       1823 ATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATG 1882
Db        461 TGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCTACAAGCGTGAAAGCCC 520

Qy       1883 TTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCAT 1942
Db        521 CTAAGTCTCCAGAGTTGCATTTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGC 580

Qy       1943 GGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGG 2002
Db        581 AGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAG 640

Qy       2003 GAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGG 2062
Db        641 GGCAGCAGGAGAA----------------------------------------------- 653

Qy       2063 ACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCC 2122
Db        654 ----TGGGCCCATTTCTTGGATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACC 709

Qy       2123 CTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTG 2182
Db        710 CCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGG 769

Qy       2183 CTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGC 2242
Db        770 CAGATCAGACTGTCAGCAC---TCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTC 826

Qy       2243 CACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTC 2302
Db        827 CTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGC 886

Qy       2303 GGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCC 2362
Db        887 ACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATC 946

Qy       2363 CCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTC 2422
Db        947 CTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGT 1006

Qy       2423 TCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGG 2482
Db       1007 TGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATC 1066

Qy       2483 ATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAA 2542
Db       1067 AGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAG 1126

Qy       2543 CACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGT 2602
Db       1127 GCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGA 1186

Qy       2603 GTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACC 2662
Db       1187 AACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGA 1246

Qy       2663 ACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAG 2722
Db       1247 AGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTT 1306

Qy       2723 AGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGG 2782
Db       1307 TGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGG 1366

Qy       2783 TGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAG 2842
Db       1367 TGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTG 1426

Qy       2843 ACTC 2846
Db       1427 AATC 1430

RESULT 13
US-09-877-730-13
; Sequence 13, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 909
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-13

  Query Match             1.9%;  Score 119;  DB 4;  Length 909;
  Best Local Similarity   48.7%;  Pred. No. 8e-21;
  Matches  416;  Conservative    1;  Mismatches  426;  Indels   12;  Gaps    3;

Qy        300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359
```

```
Db      45 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 104

Qy     360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db     105 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 164

Qy     420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db     165 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 224

Qy     480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db     225 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 284

Qy     540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db     285 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 344

Qy     597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db     345 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 404

Qy     657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db     405 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 458

Qy     717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db     459 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 518

Qy     777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db     519 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 578

Qy     837 GGATGGAAAGCCTATCTC---CACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db     579 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 638

Qy     894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db     639 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 698

Qy     954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db     699 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 758

Qy    1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db     759 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 818

Qy    1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db     819 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 878

Qy    1134 GCGCGTCAAGGTGCA 1148

Db     879 TAGAATTAAAATGTA 893
```

RESULT 14
US-09-877-730-3
; Sequence 3, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-3

Query Match 1.9%; Score 119; DB 4; Length 1143;
Best Local Similarity 48.7%; Pred. No. 9e-21;
Matches 416; Conservative 1; Mismatches 426; Indels 12; Gaps 3;

```
Qy     300 AGACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTG 359

Db     279 ATACATCAGTGAGGTGGAAGGCAGGCGAGGAGAGCAGTCCGATGAAGGATTTTATCAGTG 338

Qy     360 TCTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCAC 419

Db     339 CTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAAC 398

Qy     420 ACTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACG 479

Db     399 TATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCG 458

Qy     480 CTTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGT 539

Db     459 ATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGAC 518

Qy     540 GACCGTGCCTGAG---GAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCT 596

Db     519 AACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTA 578

Qy     597 AGATGTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCA 656

Db     579 TGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCG 638

Qy     657 ACGATTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAG 716

Db     639 ACGTAAAAGTATGGAGGCCTCGCTAACTGTG------ATTCCAGCTAAGGAGTCAAAATC 692

Qy     717 GGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAA 776

Db     693 CTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCATCAGAC 752

Qy     777 TGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA 836

Db     753 TGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCT 812

Qy     837 GGATGGAAAGCCTATCTC---CACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCAT 893

Db     813 TGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTCATGAT 872

Qy     894 CGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCAC 953

Db     873 ATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCAC 932

Qy     954 GCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCA 1013

Db     933 ACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGA 992

Qy    1014 GGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTC 1073

Db     993 ATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGA 1052

Qy    1074 CGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGG 1133

Db    1053 AGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGG 1112

Qy    1134 GCGCGTCAAGGTGCA 1148
```

```
Db       1113 TAGAATTAAAATGTA 1127

RESULT 15
US-09-877-730-25
; Sequence 25, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 1644
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-25

  Query Match             1.9%;  Score 117;  DB 4;  Length 1644;
  Best Local Similarity   48.4%;  Pred. No. 3.6e-20;
  Matches  355;  Conservative    0;  Mismatches  375;  Indels    3;  Gaps    1;

Qy       2114 AGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACG 2173
Db        470 AGGATGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATG 529

Qy       2174 GCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCC 2233
Db        530 GCTATCAGGCAGATCAGACTGTCAGCAC---TCCAGGATGCGTGTCTGTTCGTGATCGCA 586

Qy       2234 AGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCA 2293
Db        587 TGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCA 646

Qy       2294 TTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTAC 2353
Db        647 TCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCC 706

Qy       2354 GCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAG 2413
Db        707 GCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAA 766

Qy       2414 AAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCC 2473
Db        767 CTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTAC 826

Qy       2474 ACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACC 2533
Db        827 ATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAG 886

Qy       2534 GGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGT 2593
Db        887 CACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGG 946

Qy       2594 TACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACA 2653
Db        947 TTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATG 1006

Qy       2654 GCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCT 2713
Db       1007 CATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAA 1066

Qy       2714 TCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCC 2773
Db       1067 CCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCAT 1126

Qy       2774 GCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGA 2833
Db       1127 CCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGG 1186

Qy       2834 CATTCTCAGACTC 2846
Db       1187 AAACCTCTGAATC 1199

Search completed: October 15, 2003, 04:46:31
Job time : 383.754 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          October 14, 2003, 06:11:05 ; Search time 7190 Seconds
                                              (without alignments)
                                              7123.623 Million cell updates/sec

Title:           US-09-754-997A-2
Perfect score:   6604
Sequence:        1 MARADTGRGLLVLTFCLLSA..........CPLTVSPSLPRAPVSSAQVP 1252

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09754997/runat_14102003_061100_9706/app_query.fasta_1.1415
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09754997_@CGN_1_1_4823_@runat_14102003_061100_9706 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6604	100.0	3756	6	AX191303	AX191303 Sequence
	2	6604	100.0	6176	6	AX191259	AX191259 Sequence
	3	6604	100.0	6176	10	AF176694	AF176694 Mus muscu
	4	6532	98.9	6219	10	AB052620	AB052620 Mus muscu
	5	6521.5	98.8	6222	10	AB052621	AB052621 Mus muscu
	6	6351.5	96.2	6301	10	AK122535	AK122535 Mus muscu
	7	5746	87.0	6485	9	AB052622	AB052622 Homo sapi
	8	5610	84.9	3741	6	AX418406	AX418406 Sequence
	9	4903	74.2	2796	6	AX191261	AX191261 Sequence
	10	4361.5	66.0	5549	9	AB046848	AB046848 Homo sapi
c	11	3125.5	47.3	158055	2	AC110235	AC110235 Mus muscu
c	12	3064.5	46.4	224227	2	AC106509	AC106509 Rattus no
c	13	2557	38.7	176282	2	AC068507	AC068507 Homo sapi
	14	2557	38.7	198295	9	AC105129	AC105129 Homo sapi
c	15	2466.5	37.3	116580	10	AC112162	AC112162 Mus Muscu
c	16	2321	35.1	115135	10	AC112161	AC112161 Mus Muscu
	17	1747.5	26.5	3453	6	AR237556	AR237556 Sequence
	18	1746.5	26.4	3210	6	AR237553	AR237553 Sequence
	19	1746.5	26.4	3874	6	AR237568	AR237568 Sequence
	20	1691.5	25.6	2976	6	AR237558	AR237558 Sequence
	21	1691.5	25.6	3219	6	AR237561	AR237561 Sequence
	22	1472	22.3	825	6	AX191263	AX191263 Sequence
	23	1384	21.0	2715	6	AR237555	AR237555 Sequence
	24	1384	21.0	2958	6	AR237557	AR237557 Sequence
	25	1329	20.1	2481	6	AR237560	AR237560 Sequence
	26	1329	20.1	2724	6	AR237562	AR237562 Sequence
	27	1144	17.3	2139	6	AR237563	AR237563 Sequence
	28	1144	17.3	2382	6	AR237566	AR237566 Sequence
	29	1070.5	16.2	2967	9	BC042054	BC042054 Homo sapi
	30	1067	16.2	2700	6	AX497857	AX497857 Sequence
	31	1046	15.8	3198	10	BC053057	BC053057 Mus muscu
	32	1031.5	15.6	3146	10	AF026465	AF026465 Mus muscu
	33	978	14.8	4784	5	GDU07644	U07644 Gallus gall
	34	953.5	14.4	4631	9	HSU72391	U72391 Human neoge
	35	951.5	14.4	5297	6	AX587800	AX587800 Sequence
	36	951.5	14.4	5297	9	HSU61262	U61262 Human neoge
	37	947	14.3	4641	10	RNU68726	U68726 Rattus norv
	38	945.5	14.3	5414	10	MMNEOGEN	Y09535 M.musculus
	39	943	14.3	4344	6	AX054981	AX054981 Sequence
	40	943	14.3	4608	6	AR153583	AR153583 Sequence
	41	943	14.3	4608	9	HSDCCG	X76132 H.sapiens D
	42	940.5	14.2	4517	10	MMDCCTSG	X85788 M.musculus
	43	932.5	14.1	4846	5	XLU10986	U10986 Xenopus lae
	44	931	14.1	4924	10	RNU68725	U68725 Rattus norv
	45	921.5	14.0	2729	9	AK095529	AK095529 Homo sapi

ALIGNMENTS

RESULT 1

```
AX191303
LOCUS       AX191303                 3756 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 45 from Patent WO0149714.
ACCESSION   AX191303
VERSION     AX191303.1  GI:15209554
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 45 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..3756
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
BASE COUNT      774 a   1167 c   1105 g    710 t
ORIGIN

Alignment Scores:
Pred. No.:               1.74e-260      Length:         3756
Score:                   6604.00        Matches:        1252
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      6              Gaps:           0

US-09-754-997A-2 (1-1252) x AX191303 (1-3756)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy        341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy        361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy        381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy        401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy        421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy        441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy        461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy        481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy        501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy        521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy        541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy        561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy      581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy      601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy      621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy      641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy      661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy      681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy      701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy      721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy      741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy      761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy      781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy      801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy      821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy      841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy      861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy      881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy      901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy      921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy      941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy      961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy      981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy     1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy     1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy     1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy     1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy     1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy     1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy     1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy     1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy     1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy     1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy     1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy     1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy     1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
            ||||||||||||||||||||||||||||||||||||
Db     3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 2
AX191259
LOCUS       AX191259                6176 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 1 from Patent WO0149714.
ACCESSION   AX191259
VERSION     AX191259.1  GI:15209510
KEYWORDS    .
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 1 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..6176
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             1..3759
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51120.1"
                     /db_xref="GI:15209511"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

Alignment Scores:
Pred. No.:              3.04e-260      Length:          6176
Score:                  6604.00        Matches:         1252
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     6              Gaps:            0

US-09-754-997A-2 (1-1252) x AX191259 (1-6176)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy        341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy        361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy        381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy        401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy        421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy        441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380
```

```
Qy      461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy      481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy      501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy      521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy      541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy      561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy      581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy      601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy      621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy      641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy      661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy      681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy      701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy      721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy      741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy      761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy      781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy      801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy      821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy      841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy      861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy      881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy      901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy      921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy      941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy      961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy      981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy     1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy     1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy     1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy     1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy     1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy     1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy     1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy     1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy     1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy     1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy       1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756

RESULT 3
AF176694
LOCUS       AF176694                6176 bp    mRNA    linear   ROD 27-APR-2000
DEFINITION  Mus musculus neighbor of Punc e11 protein (Nope) mRNA, complete
            cds.
ACCESSION   AF176694
VERSION     AF176694.1  GI:7650185
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6176)
  AUTHORS   Salbaum,J.M. and Kappen,C.
  TITLE     Cloning and expression of nope, a new mouse gene of the
            immunoglobulin superfamily related to guidance receptors
  JOURNAL   Genomics 64 (1), 15-23 (2000)
  MEDLINE   20175427
   PUBMED   10708514
REFERENCE   2  (bases 1 to 6176)
  AUTHORS   Salbaum,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) The Neurosciences Institute, 10640 John Jay
            Hopkins Drive, San Diego, CA 92121, USA
FEATURES             Location/Qualifiers
     source          1..6176
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="band D-E1; 24.59 cR from D9Mit143"
     gene            1..6176
                     /gene="Nope"
     CDS             1..3759
                     /gene="Nope"
                     /note="immunoglobulin superfamily; related to guidance
                     receptors"
                     /codon_start=1
                     /product="neighbor of Punc e11 protein"
                     /protein_id="AAF65930.1"
                     /db_xref="GI:7650186"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
     misc_feature    1..23
                     /gene="Nope"
                     /note="unable to sequence region in mRNA"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

Alignment Scores:
Pred. No.:              3.04e-260       Length:         6176
Score:                  6604.00         Matches:        1252
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-754-997A-2 (1-1252) x AF176694 (1-6176)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780
```

```
Qy        261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy        341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy        361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy        381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy        401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy        421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy        441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy        461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy        481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy        501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy        521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy        541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy        561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy        581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy        601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy        621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy        641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy        661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy        681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy        701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy        721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy        741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy        761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy        781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy        801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy        821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy        841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy        861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy        881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy        901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy        921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy        941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy        961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy        981 SerHisArgGluAlaLeuProGlyLeuSerSerGlyThrProGlyAsnProAlaLeu 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy       1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy       1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy       1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy       1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy       1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy       1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy       1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy       1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy       1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy       1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy       1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 4
AB052620
LOCUS       AB052620                6219 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36 mRNA, complete cds.
ACCESSION   AB052620
VERSION     AB052620.1  GI:11862938
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6219)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6219
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
                     /clone="DDM36"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)-vector:pDDM36ORF"
     gene            1..6219
                     /gene="DDM36"
     CDS             50..3808
                     /gene="DDM36"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36"
                     /protein_id="BAB19278.1"
                     /db_xref="GI:11862939"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPPGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPPSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1319 a   1772 c   1652 g   1476 t
ORIGIN

Alignment Scores:
Pred. No.:               2.63e-257     Length:          6219
Score:                   6532.00       Matches:         1241
Percent Similarity:      99.44%        Conservative:    4
Best Local Similarity:   99.12%        Mismatches:      7
Query Match:             98.91%        Indels:          0
DB:                      10            Gaps:            0

US-09-754-997A-2 (1-1252) x AB052620 (1-6219)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 109

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 169

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 229
```

```
Qy     61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    230 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 289

Qy     81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 349

Qy    101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 409

Qy    121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    410 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 469

Qy    141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 529

Qy    161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 589

Qy    181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
          |||||||||||||||     ||||||||||||||||||     :::||||||||||||||
Db    590 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 649

Qy    201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 709

Qy    221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 769

Qy    241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db    770 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTA 829

Qy    261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 889

Qy    281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 949

Qy    301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
          |||||||||||||||||||||||||||||||||||||||||||||||     ||||||||
Db    950 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGAT 1009

Qy    321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCC 1069

Qy    341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1129

Qy    361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1189

Qy    381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGC 1249

Qy    401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1309

Qy    421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1369

Qy    441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1429

Qy    461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1489

Qy    481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1549

Qy    501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1609

Qy    521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1669

Qy    541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1729

Qy    561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1730 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1789

Qy    581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1790 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1849

Qy    601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1850 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1909

Qy    621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1910 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1969

Qy    641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1970 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 2029

Qy    661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
          ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db   2030 AAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2089

Qy    681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2090 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2149

Qy    701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
          ||||||||||||||||||||||||||||||||     |||||||||||||||||||||||
Db   2150 TATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTCAAC 2209

Qy    721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2210 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2269

Qy    741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2270 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2329

Qy    761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2330 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2389

Qy    781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2390 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2449

Qy        801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2450 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2509

Qy        821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2510 GCGGTACAGTCCCACGGAGTGGATATGGATGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2569

Qy        841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2570 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2629

Qy        861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2630 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2689

Qy        881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2690 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2749

Qy        901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2750 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2809

Qy        921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2810 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2869

Qy        941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
              ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db       2870 ACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2929

Qy        961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db       2930 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACGAAGC 2989

Qy        981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2990 TCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3049

Qy       1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3050 TACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3109

Qy       1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3110 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3169

Qy       1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3170 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3229

Qy       1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3230 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3289

Qy       1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3290 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3349

Qy       1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db       3350 CAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAGAAG 3409

Qy       1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3410 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3469

Qy       1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3470 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3529

Qy       1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3530 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3589

Qy       1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3590 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3649

Qy       1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3650 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3709

Qy       1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3710 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3769

Qy       1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||||||||||||||||||
Db       3770 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3805
```

```
RESULT 5
AB052621
LOCUS       AB052621                6222 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36E mRNA, complete cds.
ACCESSION   AB052621
VERSION     AB052621.1  GI:11862940
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6222)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
                     /clone="DDM36E"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)~vector:pDDM36EORF"
     gene            1..6222
                     /gene="DDM36E"
     CDS             50..3811
                     /gene="DDM36E"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36E"
                     /protein_id="BAB19279.1"
                     /db_xref="GI:11862941"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
```

```
                    VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                    AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                    QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                    PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                    GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                    LGKEEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVH
                    NQSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGG
                    RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPT
                    PDLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLV
                    TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                    SPLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLE
                    SDTRYFFKMGARTEVGPGPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLA
                    CMCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWS
                    PPPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRP
                    ALTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGC
                    PDLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEA
                    ASASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1321 a   1772 c   1653 g   1476 t
ORIGIN

Alignment Scores:
Pred. No.:              7.03e-257       Length:          6222
Score:                  6521.50         Matches:         1241
Percent Similarity:     99.36%          Conservative:    4
Best Local Similarity:  99.04%          Mismatches:      7
Query Match:            98.75%          Indels:          1
DB:                     10              Gaps:            1

US-09-754-997A-2 (1-1252) x AB052621 (1-6222)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
Db         50 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 109

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
Db        110 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 169

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
Db        170 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 229

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
Db        230 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 289

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
Db        290 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 349

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
Db        350 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 409

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
Db        410 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 469

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
Db        470 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 529

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
Db        530 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 589

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
Db        590 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 649

Qy        201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
Db        650 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 709

Qy        221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
Db        710 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 769

Qy        241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
Db        770 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTA 829

Qy        261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
Db        830 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 889

Qy        281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
Db        890 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 949

Qy        301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
Db        950 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGAT 1009

Qy        321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
Db       1010 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCC 1069

Qy        341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
Db       1070 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1129

Qy        361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
Db       1130 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1189

Qy        381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
Db       1190 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGC 1249

Qy        401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
Db       1250 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1309

Qy        421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
Db       1310 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1369

Qy        441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
Db       1370 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1429

Qy        461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
Db       1430 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1489

Qy        481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
Db       1490 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1549

Qy        501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
Db       1550 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1609

Qy        521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
Db       1610 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1669

Qy        541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
Db       1670 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1729

Qy        561 TyrLysIleGluTyrGlyLeuGlyLys---GluAspGlnValPheSerThrGluValPro 579
Db       1730 TACAAGATAGAGTACGGTTTGGGGAAGGAAGAAGATCAGGTTTTCTCCACCGAGGTGCCT 1789
```

```
Qy       580 GlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArg 599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1790 GGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGG 1849

Qy       600 IleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThr 619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1850 ATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACA 1909

Qy       620 ProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAla 639
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1910 CCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCA 1969

Qy       640 LysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGly 659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1970 AAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGA 2029

Qy       660 TyrLysLeuTyrTrpGlyGluValGlyThrGluGluAlaAspGlyAspArgProPro 679
             |||||||||||||||      |||||||||||||||||||||||||||||||||||||||
Db      2030 TACAAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCA 2089

Qy       680 GlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLys 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2090 GGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAG 2149

Qy       700 GlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPhe 719
             |||||||||||||||||||||||||||||||||||||||      |||||||||||||||
Db      2150 CAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTC 2209

Qy       720 AsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThr 739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2210 AACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACG 2269

Qy       740 ProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSer 759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2270 CCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCA 2329

Qy       760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2330 AACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATT 2389

Qy       780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2390 GTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTAC 2449

Qy       800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2450 TATACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAG 2509

Qy       820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2510 TTTGCGGTACAGTCCCACGGAGTGGATATGGATGGCCCTTTGGCTCCGTCGTAGAACGC 2569

Qy       840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2570 TCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACA 2629

Qy       860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2630 CCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAG 2689

Qy       880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2690 TATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACC 2749

Qy       900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2750 ACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTC 2809

Qy       920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2810 TTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTG 2869

Qy       940 IleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIle 959
             |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db      2870 ATTACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATC 2929

Qy       960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      2930 GTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACGA 2989

Qy       980 SerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAla 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2990 AGCTCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCG 3049

Qy      1000 LeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSer 1019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3050 CTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCC 3109

Qy      1020 LeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAla 1039
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3110 CTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCT 3169

Qy      1040 GluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLys 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3170 GAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAG 3229

Qy      1060 IleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeu 1079
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3230 ATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTG 3289

Qy      1080 ProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThr 1099
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3290 CCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACC 3349

Qy      1100 GlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLys 1119
             |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db      3350 GGGCAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAG 3409

Qy      1120 LysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPhe 1139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3410 AAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTC 3469

Qy      1140 GlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeu 1159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3470 GGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTG 3529

Qy      1160 ThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrp 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3530 ACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGG 3589

Qy      1180 LeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCys 1199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3590 CTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGC 3649

Qy      1200 LeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIle 1219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3650 TTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATA 3709

Qy      1220 GluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerPro 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3710 GAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCA 3769

Qy      1240 SerLeuProArgAlaProValSerSerAlaGlnValPro 1252
             |||||||||||||||||||||||||||||||||||||||
Db      3770 AGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3808
```

```
RESULT 6
AK122535
LOCUS       AK122535                6301 bp    mRNA    linear   ROD 15-MAR-2003
DEFINITION  Mus musculus mRNA for mKIAA1628 protein.
ACCESSION   AK122535
VERSION     AK122535.1  GI:28972801
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
            Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: II. The complete nucleotide sequences of 400 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10, 35-48 (2003)
REFERENCE   2  (bases 1 to 6301)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1..6301
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbg00013"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1..6301
                     /gene="mKIAA1628"
     CDS             <1..3887
                     /gene="mKIAA1628"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=3
                     /evidence=not_experimental
                     /product="mKIAA1628 protein"
                     /protein_id="BAC65817.1"
                     /db_xref="GI:28972802"
                     /translation="EGSQENPSAWHGRGELPLPQETTVKLSCDEGPLQVILGPEQAVV
                     LDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRI
                     WKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGL
                     PAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSARQRFSQEAS
                     LTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSWVRQGDQSK
                     KELGIQTSGCDAALQGCSGNLVVGEGVEGTVPASPSSLLLLCHQMESLSPRMSSFWAG
                     PIYSSPARSLGTLESMSAEPTSPARVISPLRLLSSECLLPQPSRRRPRRSRGRGPAPR
                     ASCAGRPGSHGPRCTGCTTGSRCDPMGRVKVQGGGGSLVITQIGLQDAGYYQCVAENS
                     AGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQK
                     ARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDV
                     PSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEEDQVFSTEVPGNE
                     TQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAK
                     MESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKV
                     KQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVH
                     AESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKP
                     FTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTE
                     PNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYFFKMGARTEVGP
                     GPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLACMCAGLRRSSHREALPG
                     LSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSPPPSDVEDKAEVHSLMG
                     GSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPALTRALLPPAGTGQTLL
                     LQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCPDLHLQDLEPEEPLTAE
                     TLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAASASCSCSDLQPSTAIE
                     EAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1362 a   1770 c   1669 g   1500 t
ORIGIN

Alignment Scores:
Pred. No.:              6e-250          Length:         6301
Score:                  6351.50         Matches:        1220
Percent Similarity:     95.40%          Conservative:   4
Best Local Similarity:  95.09%          Mismatches:     8
Query Match:            96.18%          Indels:         52
DB:                     10              Gaps:           2
```

```
US-09-754-997A-2 (1-1252) x AK122535 (1-6301)

Qy       21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       39 AGAGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 98

Qy       41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       99 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 158

Qy       61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      159 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 218

Qy       81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      219 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 278

Qy      101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      279 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 338

Qy      121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      339 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 398

Qy      141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      399 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 458

Qy      161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      459 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 518

Qy      181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
            |||||||||||||||   ||||||||||||||||||       :::||||||||||||||
Db      519 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 578

Qy      201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      579 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 638

Qy      221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      639 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 698

Qy      241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||
Db      699 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTA 758

Qy      261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGln--- 279
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      759 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACA-GGG 817

Qy      279 ------------------------------------------------------------ 279

Db      818 GGATCAGAGTAAGAAGGAACTGGGGATTCAAACCAGCGGCTGTGATGCAGCTTTGCAGGG 877

Qy      279 ------------------------------------------------------------ 279

Db      878 ATGTAGTGGGAACCTCGTAGTTGGAGAGGGTGTAGAGGGGACAGTTCCAGCGTCCCCCTC 937

Qy      280 ------------------------AspGlyLysProIleSerThrAspValIleValLeu 291
                                    ||||||||||||||||||||||||||||||||||||
Db      938 CAGCCTGCTCCTTCTCTGTCACCAGATGGAAAGCCTATCTCCACGGATGTCATCGTTCTG 997

Qy      292 GlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCys 311
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      998 GGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGC 1057

Qy      312 ArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeu 331
            |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1058 CGAGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTT 1117

Qy        332 AlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAla 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 GCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 1177

Qy        352 ArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGly 371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 1237

Qy        372 IleProLeuArgProAsn-GlyArgValLysValGlnGlyGlyGlyGlySerLeuValIl 391
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db       1238 ATCCCGTTGCGACCCAATGGGCCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCAT 1297

Qy        391 eThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGl 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 CACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGG 1357

Qy        411 yThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGlyLeuProSerAlaPr 431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 AACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCC 1417

Qy        431 oThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAlaTrpGluArgPr 451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 GACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCC 1477

Qy        451 oGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyVa 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 TGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGT 1537

Qy        471 lAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAs 491
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 GGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGA 1597

Qy        491 pLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSe 511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 CCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAG 1657

Qy        511 rArgThrSerSerProAlaLeuValHisThrLeuAspAspValProSerAlaAlaProGl 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCA 1717

Qy        531 nLeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeuProLeuProSe 551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 GCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTC 1777

Qy        551 rSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLys---Gl 570
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db       1778 CAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGA 1837

Qy        570 uAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLe 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACT 1897

Qy        590 uGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyVa 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 TCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGT 1957

Qy        610 lProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSerHisValProPh 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 CCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTT 2017

Qy        630 eAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuValValSerTrpGlnPr 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 TGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCC 2077

Qy        650 oProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGl 670
              ||||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db       2078 GCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGA 2137

Qy        670 uGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGl 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 GGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGG 2197

Qy        690 yProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyAr 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 GCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAG 2257

Qy        710 gProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTr 730
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GCTGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTG 2317

Qy        730 pLysGlyLysThrGluLysAlaProThrProAspLeuProIleGlnArgGlyProProLe 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GAAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCT 2377

Qy        750 uProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLy 770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GCCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAA 2437

Qy        770 sLysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGl 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 GAAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGG 2497

Qy        790 yLeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGl 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2498 GCTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGG 2557

Qy        810 yGlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAs 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 CGGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGA 2617

Qy        830 pGlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPr 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 TGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCC 2677

Qy        850 oSerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPr 870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 TTCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCC 2737

Qy        870 oThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGl 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 CACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCA 2797

Qy        890 nProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHi 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 GCCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCA 2857

Qy        910 sGlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPr 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 TGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGAGCCCGCACAGAGGTGGGGCC 2917

Qy        930 oGlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLe 950
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db       2918 TGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAAAGACATTCTCAGACTCCTT 2977

Qy        950 uAspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLe 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2978 GGATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCT 3037

Qy        970 uAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSe 990
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       3038 GGCCTGCATGTGTGCTGGCCTACGACGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGTC 3097

Qy        990 rSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSe 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 CTCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAG 3157

Qy       1010 rValProAlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAspTrpSerPr 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 TGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCC 3217

Qy       1030 oProProSerAspValGluAspLysAlaGluValHisSerLeuMetGlyGlySerValSe 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 ACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTC 3277
```

```
Qy       1050 rAspCysArgGlyHisSerLysArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTr 1070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 AGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTG 3337

Qy       1070 pAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuTh 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 GGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGAC 3397

Qy       1090 rArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTy 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 CCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTATA 3457

Qy       1110 rAspGlyIleLysSerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGl 1130
              ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 TGACGCCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGA 3517

Qy       1130 uAlaGluValIleValHisSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLe 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3518 AGCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCT 3577

Qy       1150 uGlnAspLeuGluProGluGluProLeuThrAlaGluThrLeuProSerThrSerGlyAl 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3578 CCAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGC 3637

Qy       1170 aValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProTh 1190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3638 TGTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAAC 3697

Qy       1190 rThrSerGlyProGluArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCy 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3698 AACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTG 3757

Qy       1210 sSerAspLeuGlnProSerThrAlaIleGluGluAlaProGlyLysSerCysGlnProLy 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3758 CTCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAA 3817

Qy       1230 sAlaLeuCysProLeuThrValSerProSerLeuProArgAlaProValSerSerAlaGl 1250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3818 AGCCCTGTGTCCTCTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCA 3877

Qy       1250 nValPro 1252
              |||||||
Db       3878 GGTCCCC 3884
```

```
RESULT 7
AB052622
LOCUS       AB052622                6485 bp    mRNA    linear   PRI 16-MAR-2002
DEFINITION  Homo sapiens hDDM36 mRNA, complete cds.
ACCESSION   AB052622
VERSION     AB052622.1  GI:19570397
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6485)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6485
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="hDDM36"
                     /cell_line="NHOst"
                     /cell_type="osteoblast"
                     /note="similar to mouse DDM36 and Nope
                     (NM_020043)~vector:phDDM36"
     gene            1..6485
                     /gene="hDDM36"
     CDS             211..3963
                     /gene="hDDM36"
                     /note="similar to mouse DDM36 and Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="hDDM36"
                     /protein_id="BAB86306.1"
                     /db_xref="GI:19570398"
                     /translation="MARGDAGRGRGLLALTFCLLAARGELLLPQETTVELSCGVGPLQ
                     VILGPEQAAVLNCSLGAAAAGPPTRVTWSKDGDTLLEHDHLHLLPNGSLWLSQPLAPN
                     GSDESVPEAVGVIEGNYSCLAHGPLGVLASQTAVVKLATLADFSLHPESQTVEENGTA
                     RFECHIEGLPAPIITWEKDQVTLPEEPRLIVLPNGVLQILDVQESDAGPYRCVATNSA
                     RQHFSQEALLSVAHRGSLASTRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVS
                     WVRQDGKPISTDVIVLGRTNLLIANAQPWHSGVYVCRANKPRTRDFATAAAELRVLAA
                     PAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNGAPLRPNGRVKVQGGGGSLVI
                     TQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPSAPTRVTATPLSSSAVLVAWE
                     RPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQ
                     LGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVAWLPLPPSLSNGQVVKYKIEY
                     GLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAGTAAGFGAPSQWMHHRTPSMH
                     NQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYKLYWREVGAEEEANGDRLPGG
                     RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPA
                     PDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFSPWGLRNASLV
                     TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                     SPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEHQWTLLTTQGNIFSAEVHGLE
                     SDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLDMHSVTGIIVGVCLGLLCLLA
                     CMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPPSPPAAHELESLVHPHPQDWS
                     PPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSGLSWAGSWAGCELPQAGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACRNQVEAEVIVHSDFSASNGNP
                     DLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDRELGGCELAAPGPDRLTCLPEAA
                     SASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLPRSPVSSSA"
BASE COUNT     1293 a   1926 c   1772 g   1494 t
ORIGIN

Alignment Scores:
Pred. No.:                2.87e-225      Length:          6485
Score:                    5746.00        Matches:         1098
Percent Similarity:       92.09%         Conservative:    54
Best Local Similarity:    87.77%         Mismatches:      96
Query Match:              87.01%         Indels:          5
DB:                       9              Gaps:            2

US-09-754-997A-2 (1-1252) x AB052622 (1-6485)

Qy          1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
              ||||||||     |||         |||||||||||||||    ||||||||||||||||||
Db        211 ATGGCGCGGGGGGACGCCGGCCGCGGCCGTGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 270

Qy         19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
              :::|||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        271 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 330

Qy         39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
              |||||||||||||||||||||||||||||||||||||||    ||||||:::|||:::||||||
Db        331 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 390

Qy         59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
              |||   |||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        391 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 450

Qy         79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
              ||||||:::::|||||||||||||||||||||||||||||||||||    ||||||
Db        451 GAGCACGACCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 510
```

```
Qy      99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     511 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 567
Qy     119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db     568 TCGTGCCTAGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 627
Qy     139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db     628 GCCACACTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 687
Qy     159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db     688 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 747
Qy     179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db     748 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 807
Qy     199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db     808 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 867
Qy     219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db     868 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACC 927
Qy     239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db     928 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 987
Qy     259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db     988 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 1047
Qy     279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db    1048 CAAGACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 1107
Qy     299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db    1108 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 1167
Qy     319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGln 338
Db    1168 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAG 1227
Qy     339 AlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSer 358
Db    1228 GCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCG 1287
Qy     359 GlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGly 378
Db    1288 GGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGG 1347
Qy     379 ArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAsp 398
Db    1348 CGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGAC 1407
Qy     399 AlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaPro 418
Db    1408 GCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCG 1467
Qy     419 LeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrPro 438
Db    1468 CTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCA 1527
Qy     439 LeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIle 458
Db    1528 CTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATC 1587
Qy     459 IleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPhe 478
Db    1588 ATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTT 1647
Qy     479 AlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyr 498
Db    1648 GCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTAT 1707
Qy     499 GluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeu 518
Db    1708 GAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTG 1767
Qy     519 ValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsn 538
Db    1768 GTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAAC 1827
Qy     539 ProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnVal 558
Db    1828 CCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTG 1887
Qy     559 LeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluVal 578
Db    1888 GTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTG 1947
Qy     579 ProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgVal 598
Db    1948 CGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTA 2007
Qy     599 ArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArg 618
Db    2008 CGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGG 2067
Qy     619 ThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArg 638
Db    2068 ACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAG 2127
Qy     639 AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSer 658
Db    2128 GCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCT 2187
Qy     659 GlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgPro 678
Db    2188 GGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTG 2247
Qy     679 ProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysVal 698
Db    2248 CCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTG 2307
Qy     699 LysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAla 718
Db    2308 AAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCT 2367
Qy     719 PheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaPro 738
Db    2368 TTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCG 2427
Qy     739 ThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGlu 758
Db    2428 GCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAA 2487
Qy     759 SerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLys 778
Db    2488 TCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAG 2547
Qy     779 IleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThr 798
Db    2548 ATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACC 2607
Qy     799 TyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyr 818
Db    2608 TATTACACCAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATAC 2667
Qy     819 GluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGlu 838
```

```
Db     2668 GAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAG 2727
Qy      839 ArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeu 858
Db     2728 CGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTG 2787
Qy      859 ThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleVal 878
Db     2788 ACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTG 2847
Qy      879 GluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeu 898
Db     2848 GAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTC 2907
Qy      899 ThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyr 918
Db     2908 ACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTAC 2967
Qy      919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
Db     2968 TTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGAT 3027
Qy      939 ValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIle 958
Db     3028 GTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATC 3087
Qy      959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
Db     3088 ATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGC 3147
Qy      979 GlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnPro 998
Db     3148 CGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCC 3207
Qy      999 AlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGlu 1018
Db     3208 GCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAG 3267
Qy     1019 SerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLys 1038
Db     3268 TCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGG 3327
Qy     1039 AlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArg 1058
Db     3328 GCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGA 3387
Qy     1059 LysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGlu 1078
Db     3388 AAGATCTCCTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAG 3447
Qy     1079 LeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGly 1098
Db     3448 CTGCCCCAG---GCAGGCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGA 3504
Qy     1099 ThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArg 1118
Db     3505 ACTGGGCAGACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGGAGG 3564
Qy     1119 LysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAsp 1138
Db     3565 AAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGAC 3624
Qy     1139 PheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluPro 1158
Db     3625 TTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCC 3684
Qy     1159 LeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAsp 1178
Db     3685 CTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCC 3744
Qy     1179 TrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThr 1198
Db     3745 TGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACC 3804
Qy     1199 CysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAla 1218
Db     3805 TGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTG 3864
Qy     1219 IleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSer 1238
Db     3865 CTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGC 3924
Qy     1239 ProSerLeuProArgAlaProValSerSerAla 1249
Db     3925 CCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3957
```

```
RESULT 8
AX418406
LOCUS       AX418406               3741 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0206329.
ACCESSION   AX418406
VERSION     AX418406.1  GI:21523336
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Rastelli,L., Shimkets,R.A., Zerhusen,B., Malyankar,U.M. and
            Padigaru,M.
  TITLE     Human polynucleotides and polypeptides encoded thereby
  JOURNAL   Patent: WO 0206329-A 1 24-JAN-2002;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3741
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
                     /note="NOVX nucleic acid with homology to NOPE-like
                     protein"
BASE COUNT      708 a   1209 c   1150 g    673 t       1 others
ORIGIN

Alignment Scores:
Pred. No.:              5.38e-220      Length:          3741
Score:                  5610.00        Matches:         1086
Percent Similarity:     91.07%         Conservative:    56
Best Local Similarity:  86.60%         Mismatches:      101
Query Match:            84.95%         Indels:          13
DB:                     6              Gaps:            8

US-09-754-997A-2 (1-1252) x AX418406 (1-3741)

Qy        1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
Db        1 ATGGCGCGGGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 60
Qy       19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
Db       61 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 120
Qy       39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
Db      121 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 180
Qy       59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
Db      181 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 240
Qy       79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      241 GAGCACGACCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 300
Qy       99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db      301 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 357
```

```
Qy   119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db   358 TCGTGCCTAGCCCACGGNCCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 417

Qy   139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db   418 GCCAGTCTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 477

Qy   159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db   478 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 537

Qy   179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db   538 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 597

Qy   199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db   598 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 657

Qy   219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db   658 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACC 717

Qy   239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db   718 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 777

Qy   259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db   778 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 837

Qy   279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db   838 ---GACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 894

Qy   299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db   895 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 954

Qy   319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgVal---LeuAlaAlaProAlaIleSer 337
Db   955 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACT 1014

Qy   338 GlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAla 357
Db  1015 CAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCG 1074

Qy   358 SerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsn 377
Db  1075 TCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAAC 1134

Qy   378 GlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGln 397
Db  1135 GGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAG 1194

Qy   398 AspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAla 417
Db  1195 GACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCG 1254

Qy   418 ProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThr 437
Db  1255 TCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACG 1314

Qy   438 ProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGln 457
Db  1315 CCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAG 1374

Qy   458 IleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGln 477
Db  1375 ATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAG 1434

Qy   478 PheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAsp 497
Db  1435 TTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGAT 1494

Qy   498 TyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAla 517
Db  1495 TATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCA 1554

Qy   518 LeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerPro 537
Db  1555 CTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCC 1614

Qy   538 AsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGln 557
Db  1615 AACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAG 1674

Qy   558 ValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGlu 577
Db  1675 GTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAG 1734

Qy   578 ValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArg 597
Db  1735 GTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGA 1794

Qy   598 ValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHis 617
Db  1795 GTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCAC 1854

Qy   618 ArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysVal 637
Db  1855 AGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTG 1914

Qy   638 ArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIle 657
Db  1915 CAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATC 1974

Qy   658 SerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArg 677
Db  1975 TCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGC 2034

Qy   678 ProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLys 697
Db  2035 CTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAA 2094

Qy   698 ValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuVal 717
Db  2095 GTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTG 2154

Qy   718 AlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAla 737
Db  2155 GCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCG 2214

Qy   738 ProThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAla 757
Db  2215 CCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCG 2274

Qy   758 GluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrVal 777
Db  2275 GAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTC 2334

Qy   778 LysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuVal 797
Db  2335 AAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTC 2394

Qy   798 ThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLys 817
Db  2395 ACCTATTAC---AGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAA 2451

Qy   818 TyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValVal 837
Db  2452 TACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTG 2511

Qy   838 GluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerPro 857
```

```
Db      2512 GAGCGCTCCACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCC 2571

Qy       858 LeuThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIle 877
Db      2572 CTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATC 2631

Qy       878 ValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeu 897
Db      2632 GTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTG 2691

Qy       898 LeuThrThr------GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAsp 915
Db      2692 CTCACCACGCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGAC 2751

Qy       916 ThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArg 935
Db      2752 ACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGC 2811

Qy       936 LeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaVal 955
Db      2812 CTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTC 2871

Qy       956 ThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAla 975
Db      2872 ACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCT 2931

Qy       976 GlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrPro 995
Db      2932 GGCCTGCGCCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCC 2991

Qy       996 GlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHis 1015
Db      2992 GGGAATCCCGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCAGCCCCCCAGCTGCCCAT 3051

Qy      1016 GluLeuGluSerLeuValHisProArgProGlnAspTrpSerProProProSerAspVal 1035
Db      3052 GAATTGGAGTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTG 3111

Qy      1036 GluAspLysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHis 1055
Db      3112 GAGGACAGGGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCAC 3171

Qy      1056 SerLysArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAla 1075
Db      3172 TCCAAAAGAAAGGTAAGT---GCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCA 3228

Qy      1076 GlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuPro 1095
Db      3229 GGCTGTGAGCTGCCCCAG---GCAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCC 3285

Qy      1096 ProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSer 1115
Db      3286 CCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGAT------CAGGGC 3339

Qy      1116 AsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleVal 1135
Db      3340 AATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTC 3399

Qy      1136 HisSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluPro 1155
Db      3400 CACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCT 3459

Qy      1156 GluGluProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGln 1175
Db      3460 GAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAG 3519

Qy      1176 GlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGlu 1195
Db      3520 GGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGAC 3579

Qy      1196 ArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnPro 1215
Db      3580 AGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCA 3639

Qy      1216 SerThrAlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeu 1235
Db      3640 GGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTA 3699

Qy      1236 ThrValSerProSerLeuProArgAlaProValSerSerAla 1249
Db      3700 GGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741
```

RESULT 9
AX191261

```
LOCUS       AX191261                2796 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 3 from Patent WO0149714.
ACCESSION   AX191261
VERSION     AX191261.1  GI:15209512
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 3 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2796
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             <1..>2796
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51121.1"
                     /db_xref="GI:15209513"
                     /translation="GELPLPQETTVKLSCDEGPLQVILGPEQAVVLDCTLGATAAGPP
                     TRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRIWKVTEGSYSCLAH
                     SPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGLPAPIITWEKDQVT
                     VPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSARQRFSQEASLTVALRGSLEATR
                     GQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSWVRQDGKPISTDVIVLGRTNLL
                     IASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAPAISQAPEALSRTRASTARFVC
                     RASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGTA
                     CAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQKARGV
                     DNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDVPSAA
                     PQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEDQVFSTEVPGNETQLTL
                     NSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAKMESLV
                     VSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKVKQYEL
                     TQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVHAESNS
                     STSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKPFTKYE
                     FAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEI
                     VEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSR
                     LQDVITLQETFSDSLDVH"
BASE COUNT      591 a    849 c    821 g    535 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2.45e-191	Length:	2796
Score:	4903.00	Matches:	932
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	74.24%	Indels:	0
DB:	6	Gaps:	0

US-09-754-997A-2 (1-1252) x AX191261 (1-2796)

```
Qy        22 GlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyProLeu 41
Db         1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy        42 GlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAla 61
Db        61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
```

```
Qy     62 AlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHisGlu 81
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy     82 AsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAsp 101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy    102 SerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCysLeu 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy    122 AlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThrLeu 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy    142 GluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArgPhe 161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy    162 GluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnValThr 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy    182 ValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspVal 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy    202 GlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPhe 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy    222 SerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGln 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy    242 AspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnValVal 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy    262 MetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAspGly 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy    282 LysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAla 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy    302 GlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPhe 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy    322 AlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGlu 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy    342 AlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluPro 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy    362 ArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLys 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy    382 ValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyr 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy    402 TyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaVal 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy    422 ValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSer 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy    442 SerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPhe 461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy    462 SerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsn 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy    482 AsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyr 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy    502 ValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThr 521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy    522 LeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAsp 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy    542 IleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyr 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy    562 LysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsn 581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy    582 GluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSer 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy    602 AlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGly 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy    622 ValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMet 641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy    642 GluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLys 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy    662 LeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGly 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy    682 ArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyr 701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy    702 GluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLys 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy    722 HisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAsp 741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy    742 LeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy        802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy        822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy        842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy        862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy        882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy        902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy        942 LeuGlnGluThrPheSerAspSerLeuAspValHis 953
              ||||||||||||||||||||||||||||||||||||
Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796

RESULT 10
AB046848
LOCUS       AB046848                5549 bp    mRNA    linear   PRI 22-FEB-2001
DEFINITION  Homo sapiens mRNA for KIAA1628 protein, partial cds.
ACCESSION   AB046848
VERSION     AB046848.1  GI:10047332
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Kikuno,R., Nakayama,M., Hirosawa,M. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XVIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 7 (4), 273-281 (2000)
  MEDLINE   20450683
   PUBMED   10997877
REFERENCE   2  (bases 1 to 5549)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..5549
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="fh14247"
                     /note="vector:pBluescriptII SK plus"
     gene            1..5549
                     /gene="KIAA1628"
     CDS             <81..3023
                     /gene="KIAA1628"
                     /note="Start codon is not identified. fh14247 cDNA clone
                     for KIAA1628 has a 40-bp different sequence replacing a
                     108-bp sequence at the position between 737 and 846, and
                     an 88-bp insertion after the position 983 of the sequence
                     of KIAA1628, respectively."
                     /codon_start=1
                     /product="KIAA1628 protein"
                     /protein_id="BAB13454.1"
                     /db_xref="GI:10047333"
                     /translation="MGRRKGRSALSRSSLERVNQGEVKVIGRGAREGGGGQAWGTGGG
                     GRDLSCGIPRSASPLAPLAAPAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNG
                     APLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPS
                     APTRVTATPLSSSAVLVAWERPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTEL
                     QVRDLEPNTDYEFYVVAYSQLGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVA
                     WLPLPPSLSNGQVVKYKIEYGLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAG
                     TAAGFGAPSQWMHHRTPSMHNQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYK
                     LYWREVGAEEEANGDRLPGGRGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAF
                     NKHEDGYAAVWKGKTEKAPAPDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTV
                     KIVNYTVRFSPWGLRNASLVTYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGS
                     VVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEH
                     QWTLLTTQGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLD
                     MHSVTGIIVGVCLGLLCLLACMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPP
                     SPPAAHELESLVHPHPQDWSPPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSG
                     LSWAGSWAGCELPQAGPRPALTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACR
                     NQVEAEVIVHSDFSASNGNPDLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDREL
                     GGCELAAPGPDRLTCLPEAASASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLP
                     RSPVSSSA"
BASE COUNT     1157 a   1546 c   1496 g   1350 t
ORIGIN

Alignment Scores:
Pred. No.:              6.04e-169       Length:         5549
Score:                  4361.50         Matches:        822
Percent Similarity:     92.55%          Conservative:   35
Best Local Similarity:  88.77%          Mismatches:     68
Query Match:            66.04%          Indels:         1
DB:                     9               Gaps:           1

US-09-754-997A-2 (1-1252) x AB046848 (1-5549)

Qy        324 AlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeu 343
              :::|||:::   |||         |||||||||||||||:::|||||||||||||||||
Db        243 TCAGCGTCGCCCCTCGCGCCCCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTG 302

Qy        344 SerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgPro 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCA 362

Qy        364 AlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGln 383
              |||||     ||||||||:::|||   ||||||||||||||||||||||||||||||||
Db        363 GCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAG 422

Qy        384 GlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGln 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAG 482

Qy        404 CysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValVal 423
              ||||||||||||||||||||||||   |||||||||||||||   |||||||||||||||
Db        483 TGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTG 542

Qy        424 ArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSer 443
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        543 CGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCT 602

Qy        444 ValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeu 463
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        603 GTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTC 662

Qy        464 HisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAsp 483
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        663 CACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAACAACGAC 722

Qy        484 ThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValVal 503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    723 ACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTACGTGGTG 782

Qy    504 AlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAsp 523

Db    783 GCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACACACTGGAT 842

Qy    524 AspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArg 543

Db    843 GATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGACATCAGG 902

Qy    544 ValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIle 563

Db    903 GTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTACAAGATA 962

Qy    564 GluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThr 583

Db    963 GAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACA 1022

Qy    584 GlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGly 603

Db   1023 CAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGT 1082

Qy    604 ThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHis 623

Db   1083 ACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCAC 1142

Qy    624 AsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSer 643

Db   1143 AACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATGGAGTCC 1202

Qy    644 LeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyr 663

Db   1203 CTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAAACTATAT 1262

Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683

Db   1263 TGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGGCCGTGGA 1322

Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703

Db   1323 GACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTATGAGCTG 1382

Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723

Db   1383 ACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAG 1442

Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743

Db   1443 GATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGACATGCCT 1502

Qy    744 IleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSerSerThr 763

Db   1503 ATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACA 1562

Qy    764 SerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsnTyrThr 783

Db   1563 TCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAACTACACT 1622

Qy    784 ValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThrSerSer 803

Db   1623 GTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACACCAGTTCT 1682

Qy    804 GlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGln 823

Db   1683 GGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCAG 1742

Qy    824 SerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThrLeuPro 843

Db   1743 TCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCCT 1802

Qy    844 AspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSerThrVal 863

Db   1803 GACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTCCACGGTT 1862

Qy    864 ArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeuIleLeu 883

Db   1863 CGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCTGATCCTG 1922

Qy    884 TyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsn 903

Db   1923 TACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCAGGGAAAC 1982

Qy    904 IlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGly 923

Db   1983 ATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGG 2042

Qy    924 AlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGln 943

Db   2043 GCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAG 2102

Qy    944 GluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleValGlyValCys 963

Db   2103 GAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGC 2162

Qy    964 LeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArg 983

Db   2163 CTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAGG 2222

Qy    984 GluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArg 1003

Db   2223 GAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGA 2282

Qy   1004 AlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisPro 1023

Db   2283 GCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCC 2342

Qy   1024 ArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSer 1043

Db   2343 CATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGC 2402

Qy   1044 LeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIleSerTrpAla 1063

Db   2403 CTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGATCTCCTGGGCT 2462

Qy   1064 GlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySer 1083

Db   2463 CAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAG---GCA 2519

Qy   1084 GlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeu 1103

Db   2520 GGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTG 2579

Qy   1104 LeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLysProSerPro 1123

Db   2580 TTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGGAGGAAGAAGTCACCCCCA 2639

Qy   1124 AlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGlyAlaSerLys 1143

Db   2640 GCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAAC 2699

Qy   1144 GlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThrAlaGluThr 1163

Db   2700 GGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCT 2759

Qy   1164 LeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGlu 1183

Db   2760 CCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAG 2819

Qy   1184 LeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeuProGluAla 1203

Db   2820 TTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCA 2879

Qy   1204 AlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGluGluAlaPro 1223

Db   2880 GCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCT 2939
```

```
Qy      1224 GlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSerLeuProArg 1243
             |||   ||||||||  |||:::  ||||||||     ||||||  ||||||||
Db      2940 GGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCTGCCCAGA 2999

Qy      1244 AlaProValSerSerAla 1249
             :::||||||||||||:::
Db      3000 TCCCCGGTCTCCTCCTCT 3017

RESULT 11
AC110235/c
LOCUS       AC110235               158055 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-100M12, WORKING DRAFT SEQUENCE, 17 ordered
            pieces.
ACCESSION   AC110235
VERSION     AC110235.5  GI:20800205
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-100M12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 15, 2002 this sequence version replaced gi:20531900.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22383
                Center clone name: 100_M_12
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 153807 bases at least Q40
                Consensus quality: 155134 bases at least Q30
                Consensus quality: 155784 bases at least Q20
                Insert size: 218000; agarose-fp
                Insert size: 156455; sum-of-contigs
                Quality coverage: 7.1 in Q20 bases; agarose-fp
                Quality coverage: 10.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     6459: contig of 6459 bp in length
            *     6460     6559: gap of 100 bp
            *     6560     6667: contig of 108 bp in length
            *     6668     6767: gap of 100 bp
            *     6768     7863: contig of 1096 bp in length
            *     7864     7963: gap of 100 bp
            *     7964     9426: contig of 1463 bp in length
            *     9427     9526: gap of 100 bp
            *     9527    11011: contig of 1485 bp in length
            *    11012    11111: gap of 100 bp
            *    11112    12741: contig of 1630 bp in length
            *    12742    12841: gap of 100 bp
            *    12842    14789: contig of 1948 bp in length
            *    14790    14889: gap of 100 bp
            *    14890    17826: contig of 2937 bp in length
            *    17827    17926: gap of 100 bp
            *    17927    21244: contig of 3318 bp in length
            *    21245    21344: gap of 100 bp
            *    21345    27413: contig of 6069 bp in length
            *    27414    27513: gap of 100 bp
            *    27514    36381: contig of 8868 bp in length
            *    36382    36481: gap of 100 bp
            *    36482    45104: contig of 8623 bp in length
            *    45105    45204: gap of 100 bp
            *    45205    58369: contig of 13165 bp in length
            *    58370    58469: gap of 100 bp
            *    58470    85782: contig of 27313 bp in length
            *    85783    85882: gap of 100 bp
            *    85883   112706: contig of 26824 bp in length
            *   112707   112806: gap of 100 bp
            *   112807   153927: contig of 41121 bp in length
            *   153928   154027: gap of 100 bp
            *   154028   158055: contig of 4028 bp in length.
FEATURES             Location/Qualifiers
     source          1..158055
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
```

```
                    /db_xref="taxon:10090"
                    /clone="RP23-100M12"
                    /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1. .6459
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   6560. .6667
                    /note="assembly_fragment"
     misc_feature   6768. .7863
                    /note="assembly_fragment"
     misc_feature   7964. .9426
                    /note="assembly_fragment"
     misc_feature   9527. .11011
                    /note="assembly_fragment"
     misc_feature   11112. .12741
                    /note="assembly_fragment"
     misc_feature   12842. .14789
                    /note="assembly_fragment"
     misc_feature   14890. .17826
                    /note="assembly_fragment"
     misc_feature   17927. .21244
                    /note="assembly_fragment"
     misc_feature   21345. .27413
                    /note="assembly_fragment"
     misc_feature   27514. .36381
                    /note="assembly_fragment"
     misc_feature   36482. .45104
                    /note="assembly_fragment"
     misc_feature   45205. .58369
                    /note="assembly_fragment"
     misc_feature   58470. .85782
                    /note="assembly_fragment"
     misc_feature   85883. .112706
                    /note="assembly_fragment"
     misc_feature   112807. .153927
                    /note="assembly_fragment"
     misc_feature   154028. .158055
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    41469 a  36830 c  36521 g  41635 t   1600 others
ORIGIN

Alignment Scores:
Pred. No.:               5.71e-117      Length:        158055
Score:                   3125.50        Matches:       913
Percent Similarity:      23.34%         Conservative:  2
Best Local Similarity:   23.29%         Mismatches:    6
Query Match:             47.33%         Indels:        3002
DB:                      2              Gaps:          13

US-09-754-997A-2 (1-1252) x AC110235 (1-158055)

Qy       332 AlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAla 351
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85019 GCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 84960

Qy       352 ArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGly 371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84959 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 84900

Qy       372 IleProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIle 391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84899 ATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATC 84840

Qy       392 ThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGly 411
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84839 ACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGA 84780

Qy       412 ThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGlyLeuProSerAlaPro 431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84779 ACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCG 84720

Qy       432 ThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAlaTrpGluArgPro 451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84719 ACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCT 84660

Qy       452 GluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyVal 471
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84659 GAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGTAT- 84601

Qy       471 ------------------------------------------------------------ 471

Db     84600 GTCTACTCTTGACCGCGGGAATCCCAGTTAGGGCAAGTATGGGATGGATAATTAAAGATG 84541

Qy       471 ------------------------------------------------------------ 471

Db     84540 AGAGGAAGGGATGGGGAGCAATACTGCTGGGAAGTGATCTAGTATTAGAGAATGGAGTGT 84481

Qy       471 ------------------------------------------------------------ 471

Db     84480 ACACGCACTCCTTGACGGAGAGAATGCTTCTCTTTCCAGGGTTAAAGGAGCCATCTTCCA 84421

Qy       471 ------------------------------------------------------------ 471

Db     84420 CCCCCAGCTAACTAACTGCAGGACGGGCTCCTTCAATTAGCACGGGGAGCTTCCTGGCCT 84361

Qy       471 ------------------------------------------------------------ 471

Db     84360 ATTATGGAATCAATGCAATAAGTAGACTAGATGGCAGAGGCATTTCTATCCTTCACCCTA 84301

Qy       471 ------------------------------------------------------------ 471

Db     84300 GATTGTCTGTAACAATTTAGCTATACATAGACCAGTAGGATCCCCAGACAGTCTCCAGCT 84241

Qy       471 ------------------------------------------------------------ 471

Db     84240 AGTCAGACATGGAGTGGGCTGTGACAGTCCCTTTGTCCCCCCCGTGCTACCACCTCAGGA 84181

Qy       472 ---AspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArg 490
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84180 GTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGG 84121

Qy       491 AspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAla 510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84120 GACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCC 84061

Qy       511 SerArgThrSerSerProAlaLeuValHisThrLeuAspAsp------------------ 524
             ||||||||||||||||||||||||||||||||||||||||||
Db     84060 AGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGGTAGGACCTCTGAACCT 84001

Qy       524 ------------------------------------------------------------ 524

Db     84000 GCAGTGACCTGGGACCCAGAGACACTCCCACTCAGGGACAGTGTGGCCCAAGGCCCTCTT 83941

Qy       524 ------------------------------------------------------------ 524

Db     83940 CCCTGCCTTGTGATGTCTGTTGTGCCCTATAAAGCAGTGGGCAGAGGTTCTTCTTCCCAG 83881

Qy       524 ------------------------------------------------------------ 524

Db     83880 TGTAAAATGGGGAAACTGGGGCTCAGAAAATGAAAATGACTTGCCAAGCTCAGCAGGGCA 83821

Qy       524 ------------------------------------------------------------ 524

Db     83820 GCTGGGACTTGAGCCCAGCTCTGAGCCCCCATACCATGCAGCCCTTCTTCCTAACATACA 83761

Qy       524 ------------------------------------------------------------ 524

Db     83760 TTTACTACCCAGCCATTGGATATGAAAGCTTCTGGCCCACACTGTCCACTAGCCTTGGGA 83701

Qy       524 ------------------------------------------------------------ 524

Db     83700 TAGAAAGGATGTCTCCGCATCCTCCAGCGTAGGGCTCCCCTGAGAGGCTCCCTCCCTCCC 83641
```

```
Qy       524 ------------------------------------------------------------ 524
Db     83640 TCCCTCAGATCCTGGGCCTGGGCCTACTTCTGACTTGCTGTGTGACCTGGGCAAGTCTCT 83581
Qy       524 ------------------------------------------------------------ 524
Db     83580 GATCACTCCTGACTGTGTCTTATCTTCCCTGTTGGTGGAGCGGAGCTCACGCCAATGTTC 83521
Qy       525 ----------ValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAs 541
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83520 TTCTGTCTCAGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGA 83461
Qy       541 pIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTy 561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83460 CATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTA 83401
Qy       561 rLysIleGluTyrGlyLeuGlyLys----------------------------------- 569
             ||||||||||||||||||||||||
Db     83400 CAAGATAGAGTACGGTTTGGGGAAGGAAGGTGAGTGGGGCCGGGGAGGTACAGAGATGTC 83341
Qy       569 ------------------------------------------------------------ 569
Db     83340 AGTGGGAATGGGGGTCCTCCCCTAGATGAGGGACGTTGTCAGTGGAAGTGGAGCCTAATT 83281
Qy       569 ------------------------------------------------------------ 569
Db     83280 GGGGCATGGGATGAGAAAGCTCTGGAAAAGAGCATTGGGAGACTAGATAGCCCAATGTGT 83221
Qy       569 ------------------------------------------------------------ 569
Db     83220 AAAGTATGTGTGCAAGCCCAAAGGCCTGAGTTCAATCTCAGACTTCATGTTAAAAAAAAA 83161
Qy       569 ------------------------------------------------------------ 569
Db     83160 AAAGCCTGGCACAGTGGTAGGTGCCTGCTTGTAATCCCAGAGCAGGGGAGGTGGAGACAC 83101
Qy       569 ------------------------------------------------------------ 569
Db     83100 GGGGACACCTTGAGCTCACTGGGCTAAACCAGTGAGTTCCAGGCCAGCGAGAGTCCCTGT 83041
Qy       569 ------------------------------------------------------------ 569
Db     83040 GCAGAAGAGAAGGTGGACAATACCAGAGAAATGACAACAGAGGTTGTCATCTGACCTCCA 82981
Qy       569 ------------------------------------------------------------ 569
Db     82980 CATGTATGTGTACACACACCTGCACATAGGTGAACACTCACACTCATGGACACACACGGA 82921
Qy       569 ------------------------------------------------------------ 569
Db     82920 CACAGGGTTTCGGGGCAGTTGAGGAAAAGGGAAAGCAAAGAGGGTCAGTAGCCTGCGCAC 82861
Qy       569 ------------------------------------------------------------ 569
Db     82860 AGATAAGGAGGGTGATCAGTGGGACCTTGGACTGTTCCCTTTCTGTGAAGGGAGCAATTT 82801
Qy       569 ------------------------------------------------------------ 569
Db     82800 ACCCAGTGGATCCTTGGAACCCTTCTAGGTTTTTGTGTTTGTTTATTTTTTAAGATGTAT 82741
Qy       569 ------------------------------------------------------------ 569
Db     82740 TTATTTATTTTATGTATATGAGTTCTCTATCTGCACGCCCACCTGCATGCCAGAAGAGGG 82681
Qy       569 ------------------------------------------------------------ 569
Db     82680 AATCAGATCCCAGTACAGATGGTTGTGAGCCACAGTGTGGTTGCTGGGAATTGAACTCAG 82621
Qy       569 ------------------------------------------------------------ 569
Db     82620 GACCTCTGGAAGAGCAGTCAGTGCTCCTAACAATTGAGCCATCTCTCCAGCCCCAGAAAC 82561
Qy       569 ------------------------------------------------------------ 569

Db     82560 CTTCTGGTTTACACACCAGAGAAGATGTGGGAGGCTGCAGGAGTGCACATGAGACATGGG 82501
Qy       569 ------------------------------------------------------------ 569
Db     82500 CTGCTCCTACTGGGGAGATGATGCTGGTCTGGGGGAGGTGATGTGGTTGGGAAAGGAATG 82441
Qy       569 ------------------------------------------------------------ 569
Db     82440 GAGAGTGAGGAAAAGAAAAGTGGTGAGGGTCAGGGTAAAGCTAAAGAGAGAGGCAGGAGG 82381
Qy       569 ------------------------------------------------------------ 569
Db     82380 CTGATGGGAGACAGCAGTCAGGGCTCCTGACACCAGGGACTGCTTAGTGTCCTTAGCCAG 82321
Qy       569 ------------------------------------------------------------ 569
Db     82320 CCATCAAGCCTGTGTAAGAGGTGTGTACTGGCTGTTGCTGTTGGTTCACAGGGAGGAAAA 82261
Qy       569 ------------------------------------------------------------ 569
Db     82260 GTTCATGCCTTGGTTTCTCCATATGGAACGGAGGCCAATGGGCTTCAGCAGCGATAGGTA 82201
Qy       569 ------------------------------------------------------------ 569
Db     82200 GAAGCCCAGGGTCAACCTGGAGGCATGAAACCTGCCTCCGTGCTGAGGGGTCAGTGCCCA 82141
Qy       569 ------------------------------------------------------------ 569
Db     82140 GAGGACCCTGGGGACCTGAAATGGAAAAGTCAATGTGATAAATTTAAAGATTGATGGTCC 82081
Qy       570 ----------------------------------------GluAspGlnValPheSerTh 576
                                                     |||||||||||||||||||
Db     82080 CCCAATGAAGGACACCCAGCAGCCTATCCTTTTTGTCCCAGAAGATCAGGTTTTCTCCAC 82021
Qy       576 rGluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTy 596
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82020 CGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTA 81961
Qy       596 rArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGl 616
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81960 CCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCA 81901
Qy       616 nHisArgThrProGlyValHisAsnGlnSerHis-------------------------- 627
             |||||||||||||||||||||||||||||||||
Db     81900 GCACAGGACACCTGGTGTGCACAACCAGAGCCATGGTATGGACCTACACGGGCAGACTGG 81841
Qy       627 ------------------------------------------------------------ 627
Db     81840 ACAGCGGGGACAGGTGCTCTGTGGCCTTAGATATTGCCCCTGTTATTGCCTGTTATTGCC 81781
Qy       627 ------------------------------------------------------------ 627
Db     81780 ACCCCCCCCCCCCCACAAACACACACTTGATGCACCCGGAAGGCCTGGGTGTAGGAGTGA 81721
Qy       627 ------------------------------------------------------------ 627
Db     81720 AGGGCTGGGCTCAAATATCTCCTTGTTCACTCTGTGACTTTGACAAGCCACTCGCATCTC 81661
Qy       627 ------------------------------------------------------------ 627
Db     81660 CAGAGCTTCGGTACTACCAAGGATGATAGCAGGACTAGTCATTGTAATTATACACAGAGT 81601
Qy       627 ------------------------------------------------------------ 627
Db     81600 GGTCGGCATAGAACAGGAGCTCCTTGGGTTGAGACCACGGCTCACTTGACAGAGTGCTTG 81541
Qy       627 ------------------------------------------------------------ 627
Db     81540 CCTAGCATGTACAAAGCACTAAATTCAATCCCCAGGACCACATAAACCCAGCATACCTCT 81481
Qy       627 ------------------------------------------------------------ 627
```

```
Db     81480 GATCCTGGCAGTTCAAGGTCTGCCTCACCTACCTAGTGAGTTGGGGACCAGCCTAGGAGG 81421
Qy       627 ------------------------------------------------------------ 627
Db     81420 TATGAAACTCTGTATCAAAGAAAGGAAGGAACCCATTAAAAGCTAGCTGGCTTTGCCAAG 81361
Qy       627 ------------------------------------------------------------ 627
Db     81360 AGGCCCAAACAGTAGAGTGGTTAAATGGCTACAAAGTCTCCAGGCAGGTGCTAGTGGAGG 81301
Qy       627 ------------------------------------------------------------ 627
Db     81300 CTCATGCCTTTAATCCCAGCACTCGAGAGGTGGAGGCAGGCAGATTTCTGAATTTGAAGA 81241
Qy       627 ------------------------------------------------------------ 627
Db     81240 CAGCCTGGTCTACACCAAGGTTACACAGAGAAACCCTGTCTCAAAAAACCAGAAAAGAAA 81181
Qy       627 ------------------------------------------------------------ 627
Db     81180 AAAGACTACAAAGTCAGCTACCATATTCAGACATTAGACAAGCTATAGCTTAAAAAAAAA 81121
Qy       627 ------------------------------------------------------------ 627
Db     81120 AAAAAAAAAAAAAGCTGGATTATGGAATTATAAGAATGGGCCGGGCGCGGTGGCGCACGC 81061
Qy       627 ------------------------------------------------------------ 627
Db     81060 ATGCCTTTGATCCCAGTACTCGGGAGGCAGAGGCAGGCGGATTTCTGAGTTCGAGCCTGG 81001
Qy       627 ------------------------------------------------------------ 627
Db     81000 TCTACAAAGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTCGAAAAAAG 80941
Qy       627 ------------------------------------------------------------ 627
Db     80940 AAAAAGAAAAAGAAAAAAAAATGAATGAAGTAAAAATATACAAAGGCTGCCAAGGTCGCC 80881
Qy       627 ------------------------------------------------------------ 627
Db     80880 CAGCTAATAAAGGGGCTTGCTCCATTTTCTGGGGATCTGACTTCCAATGTCTCCCCAGAA 80821
Qy       627 ------------------------------------------------------------ 627
Db     80820 TTTATATGATGAGTGAGTCTACAGCTTGCGGCCTCTGACCGCCACCAGCACGCACATAAA 80761
Qy       627 ------------------------------------------------------------ 627
Db     80760 TTAAATAAATACTTTAAATTTTTTAAAATACATATACAAAATGTTTGAAATGGGGTCTGA 80701
Qy       627 ------------------------------------------------------------ 627
Db     80700 CACACAGCCAACCCTCTACTCTTGGGCTCCGAGTGTTGCTCGTCGTGTTGATGGTCCCAA 80641
Qy       627 ------------------------------------------------------------ 627
Db     80640 ATCTCCCTCATGATCTCAGCAGCCCCTTGACCATACTAACTCTTTCAAGAGACCTGCAGA 80581
Qy       627 ------------------------------------------------------------ 627
Db     80580 AGCCCCTCTGCCCCTGACCGCGGGCCTGCGCAGACACGCCCCCTGCTGTCTCTTGTGCGT 80521
Qy       627 ------------------------------------------------------------ 627
Db     80520 GCCCGGCAGACGCGCCTGCGCAGCTAACATGCAGAGCTCAAACCCACCGGGAAGTGGCAG 80461
Qy       627 ------------------------------------------------------------ 627
Db     80460 AGTGGAGCAGTTAGGGAAAGCCCAGGCACAGCTCTATGCTGTGCTCCGCCCCTGCCCTGC 80401
Qy       627 ------------------------------------------------------------ 627
Db     80400 ACCGGCTCACCTCCCTGACCACTTCACTCCTCATTCTCTTCCCCTTGTCCCCATCATCTT 80341

Qy       628 -------------ValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSe 643
                          |||||||||||||||||||||||||||||||||||||||||||||||
Db     80340 CCTCTGCCCCCCAGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTC 80281

Qy       643 rLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTy 663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80280 CCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTA 80221

Qy       663 rTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGl 683
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80220 CTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGG 80161

Qy       683 yAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLe 703
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80160 AGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACT 80101

Qy       703 uThrGlnLeu-------------------------------------------------- 706
             ||||||||||
Db     80100 GACCCAGTTAGGTGAGTAGGCTTGGGCTTGGGAGGGTCATGGGCAAGCTAAGCACACGGA 80041

Qy       707 -------------------------ValProGlyArgProTyrGluValLysLeuValAl 718
                                      |||||||||||||  ||||||||||||||||||||
Db     80040 GAGTGACACTCATGGCCTTTCCCCAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGC 79981

Qy       718 aPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaPr 738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79980 TTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCC 79921

Qy       738 oThrPro----------------------------------------------------- 740
             |||||||
Db     79920 CACGCCAGGTGAGGGGGAGGCAGGCGGGGGTCTTGGAGCCTTCTTTTCCTTTTTCTTTAT 79861
Qy       740 ------------------------------------------------------------ 740
Db     79860 CAGCCAGCAAGCCGTGTGCATGCCACAGCCACCCGGCCCACCCTCCTGGGCAGAAATCCT 79801
Qy       740 ------------------------------------------------------------ 740
Db     79800 CCTGCTCTAGACATGCCCCTCATTCTCAAGCTAAGCAGAGCGAACGGGAGAGATTGGAAA 79741
Qy       740 ------------------------------------------------------------ 740
Db     79740 GAGAAACTCCGTGTCCGATGCGTGATTCCTGGCTCAAGCTTTAACCTTTAAACAGCTTTT 79681
Qy       740 ------------------------------------------------------------ 740
Db     79680 TAGATCGCATCCACGTATACCACATGGTCCGTCAGATTATCCCATAATTCTTTTTTTTAA 79621
Qy       740 ------------------------------------------------------------ 740
Db     79620 AGACTTACTTACTTATTTTATGTATATGAGTACACTGTAGCTGTCTTCAGACACACCAGA 79561
Qy       740 ------------------------------------------------------------ 740
Db     79560 AGAGGGCACCAGATCCCATTACAGATGGTTGTGGCTTCTGGGAATTGAACTCCAGACCTC 79501
Qy       740 ------------------------------------------------------------ 740
Db     79500 TAGAAAAGCAGTCAGTGCTCTCAACCACTGAGCCATCTCTCCAGCCCCACCCCATAATTC 79441
Qy       740 ------------------------------------------------------------ 740
Db     79440 TTAAGCAGGGGCAAGCACAACACAGTGGTTTGTCACCCTGACGATCCAGCCTGACAGTCA 79381
Qy       740 ------------------------------------------------------------ 740
Db     79380 GAAGAAACCTGGGTTCTAATTCCAGCGAGTTCAGTTTTTTCATTGAGCAAATATTTTACC 79321
Qy       740 ------------------------------------------------------------ 740
Db     79320 TAAGTATCTACCACATGACAGCTCTGTTCTGGAAACTTAGTAATGAACAACATGAATAAG 79261
```

```
Qy        740 ------------------------------------------------------------ 740
Db      79260 GTATCTGCTCTCTTGGAGCTTATTTTATGCTTATCTAATAAACATTTGTCATCACCAAAT 79201
Qy        740 ------------------------------------------------------------ 740
Db      79200 AGTATACACTGTCCCACAGTCTTCTACCTGCTTTGTAAGCAGCCTCCTAAGGACACTCAT 79141
Qy        740 ------------------------------------------------------------ 740
Db      79140 GAGGGGACAGTGAGATGGTTCAGCAGGGAAGGTGATTGCTACAAGCCACATGACCTGAGC 79081
Qy        740 ------------------------------------------------------------ 740
Db      79080 TTGGTTTCTGCACCCCCCCCCAAATTGTCCTCTGCTCCCCCACATGCACACAGCAGCACA 79021
Qy        740 ------------------------------------------------------------ 740
Db      79020 CTCCCTCCCTTCCCTCCACACAATAAATGTGTGACTGAGAAGATGCAGGAGTTAAGAATT 78961
Qy        740 ------------------------------------------------------------ 740
Db      78960 TTAAATATGTTTTAAAGATGTATTTATGTATTTATATGTGTGAGTACACTGTCACTGTTG 78901
Qy        740 ------------------------------------------------------------ 740
Db      78900 CTATCTTCAGACACAGCAGAAGAGGTCACTAGATCCCATTATAGATGGTTGTGAGCCACC 78841
Qy        740 ------------------------------------------------------------ 740
Db      78840 ATGTGGTTGCTGGGAATTGAACTCAGGGACTCTGGAAGAGCAGTCAGTGCTTAACCGCTG 78781
Qy        740 ------------------------------------------------------------ 740
Db      78780 AGCCCTCTCTTCAGCCCCCAGAATTTTAAATATTTTAAAGAAGTAAATATAAAGAAAAGA 78721
Qy        740 ------------------------------------------------------------ 740
Db      78720 AGAAAAATGTCAAGATGTAGGTATTGTCATCAGATCCTTATAAGAATAGAAAATGATGGC 78661
Qy        740 ------------------------------------------------------------ 740
Db      78660 ACAGAGAGGTTAAACAACTTCTCTAAAGCCACATACATAGCTAGGAAGTGGCAGCAATGG 78601
Qy        740 ------------------------------------------------------------ 740
Db      78600 AGCCTGGCTCCAGAGTCAGGGCATGCTTTCTCTGTCACATTCTCATCAGCCATGGGATAT 78541
Qy        740 ------------------------------------------------------------ 740
Db      78540 GTCACAGTCACTCTATGTGTGACCTAAGCAAAAAAAAATCACTTCCATTCTTGGGAGCCT 78481
Qy        740 ------------------------------------------------------------ 740
Db      78480 CAGCTGCAGTAGGAATAATGTCACCCAACTCACAGGATCACAAGAGAGGCCCGGTGCCAG 78421
Qy        740 ------------------------------------------------------------ 740
Db      78420 CTGTGGTATCACACAGCTGTAACCCCTGTGCTAAAGGACCAGAGGCACAAGGATGGAGAA 78361
Qy        740 ------------------------------------------------------------ 740
Db      78360 TTACAGGCCAGCCTAGGCTACTACATAATGAGATTCTGCCTCAAAAATAAGAGGGGGAAA 78301
Qy        740 ------------------------------------------------------------ 740
Db      78300 GGCCAAATGAAACCTACTGAGACCATGTACGTCCACCTACTCTCTGCACACTTGGGTTTC 78241
Qy        741 --------AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaG 758
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78240 CCCCTCTAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAG 78181
Qy        758 luSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValL 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78180 AGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCA 78121
Qy        778 ysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValT 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78120 AGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCA 78061
Qy        798 hrTyrTyrThr------------------------------------------------- 801
              |||||||||||
Db      78060 CCTACTATACCAGGTATCAGTGAGGGGGGAGTGTAGCTGGGGTGGTAGACTAGCGGGTTC 78001
Qy        801 ------------------------------------------------------------ 801
Db      78000 AGCTGTGGTGAGGCCACACCCAGGATCTACACAGATGTCTTTTCCGTGACTGCTGACCTG 77941
Qy        801 ------------------------------------------------------------ 801
Db      77940 GGGGGTAATAATGAGGGATGGGATTCCCCCATCCCTCAGTGGCAGAATGTGTGGTCAGCC 77881
Qy        801 ------------------------------------------------------------ 801
Db      77880 TTTGGTGAGAGTGGATGCAAGACTAGAGGCTTCCATCCGTTGGCTGGCACCCTGCCGGCC 77821
Qy        801 ------------------------------------------------------------ 801
Db      77820 CATCATTCGGAAGATAGCTTAAGGAGCTGCTGTTCTGATTGGGAAGACACCCTGTGAAAG 77761
Qy        801 ------------------------------------------------------------ 801
Db      77760 CCAACAGAGAGAGAAACTAAGCTGAGGATAGACACTTCAGGAGCGATTGGTCCCCCAGGT 77701
Qy        801 ------------------------------------------------------------ 801
Db      77700 GCAGAATAACTTAGATTAATTCGGGAAAGCGTCCCAGAGGAGAGAAGGGCTTGCTTTGGA 77641
Qy        801 ------------------------------------------------------------ 801
Db      77640 GATGAGACATGGAATAACATATGTCCGGCCTTTCTGTGGTTTGCCTGGCTGTGTGTGGGA 77581
Qy        801 ------------------------------------------------------------ 801
Db      77580 GGGGAGGTTGTGGGCCCAACAAGCAGCTCCCTCCAGGGTGGGGCCTGGCTGGAGGCAGGG 77521
Qy        801 ------------------------------------------------------------ 801
Db      77520 GCTTCTGTCTGGGGTTCTGCCTGGCTTGAGGCCTGTCTGAGGGCACTAGTCCTATCCCTT 77461
Qy        802 --------SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrG 819
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77460 CTCCTGGAAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACG 77401
Qy        819 luPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluA 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77400 AGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAAC 77341
Qy        839 rgSerThrLeuProAspArg---------------------------------------- 845
              ||||||||||||||||||||
Db      77340 GCTCCACCCTGCCTGACCG-TGAGTATCCACTGCCCGCTCCCCACCTGTGCCACCATTTG 77282
Qy        845 ------------------------------------------------------------ 845
Db      77281 CTGGGTCTGAGGAAGGAAGAAGAAGAGGAGGCTGTGCCCTCATCCCCTTAGTTAGGGCTG 77222
Qy        846 ------------------------------------ProSerThrProProSerAspL 853
                                                  ||||||||||||||||||||||||
Db      77221 TGTCTCTGACTCTCCACCCCTTCCCCTGTGCTGCAGGGCCTTCAACACCTCCTTCTGACC 77162
Qy        853 euArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGluP 873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77161 TGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGC 77102
Qy        873 roAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluH 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     77101 CCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCCGAAC 77042

Qy       893 isGlnTrpThrLeuLeuThrThrGlu---------------------------------- 901
             ||||||||||||||||||||||||||
Db     77041 ACCAGTGGACACTGCTCACCACAGA-GGGTAAGGACGCCCACCCCTGCCTCCTTGCCCTT 76983

Qy       901 ------------------------------------------------------------ 901

Db     76982 TTGTTCCAAGGGTCTCTGGGATTGGAGGCTGAAGCTACTGTCTGGGTGCAAAGTTGGTTT 76923

Qy       902 -----------------------------------------------------GlyAsnI 904
                                                                  |||||||
Db     76922 GCCCATGGTGCTTTGGCCTCCTGCCACCCTGTCTCCTCCCACTTCCCACCACAGGAAACA 76863

Qy       904 lePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGlyA 924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76862 TCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAG 76803

Qy       924 laArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGlnG 944
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db     76802 CCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAA 76743

Qy       944 luThrPheSer------------------------------------------------- 947
             ::|||||||||
Db     76742 AGACATTCTCAGGTACCAGGCAGGAGGGAGAAGGCCCTTGGGACAACTGGAGGCCCCAAG 76683

Qy       947 ------------------------------------------------------------ 947

Db     76682 ACCCCTGGATTGGCTCCCAAGGCCCTGCCTTCTGGAAATTACATTTTCTTTACTTTGGAA 76623

Qy       947 ------------------------------------------------------------ 947

Db     76622 CCCTAAGGTGTGGGGGATGGAGAAGCAGTGACAGCTTGATAGAGGGACTGTCATAGCTTT 76563

Qy       947 ------------------------------------------------------------ 947

Db     76562 GAGGGGTCTTTTATTATTTTATTCTCTCTGTGTGTGTTTGGCACAGTGCATATGCGAAGG 76503

Qy       947 ------------------------------------------------------------ 947

Db     76502 TCAGAGGACATCAGAGAGTTGGTTTTGTGGGGATCTTGCTTTGCCTGGAGGGTTGGGTGA 76443

Qy       947 ------------------------------------------------------------ 947

Db     76442 GCTCCTGAGTTCTGTGAAGAAGTGGTCAATAGGCTGGATAGACACTCACAAGGGCTCAAT 76383

Qy       947 ------------------------------------------------------------ 947

Db     76382 TAAGATACAAAATGAGATCCCAGGGCTCCTTATTGTAAATTATTGAAGATGTGAAGATGA 76323

Qy       947 ------------------------------------------------------------ 947

Db     76322 GCCAGTGGGGCATTAGCCACACATAGGACTGAGGAGCATGAAGCCCTGAGTCCAAATGCA 76263

Qy       947 ------------------------------------------------------------ 947

Db     76262 CACACACACACACACACACACACACACACACACACACCGTAAAACTTGGCTGTGGCTGCT 76203

Qy       947 ------------------------------------------------------------ 947

Db     76202 GTAATCCTTGCTGAGGATGCAGAGACAGGATATTTCTGGGGCTGAGAGGGACTGGGGAAG 76143

Qy       947 ------------------------------------------------------------ 947

Db     76142 TCACGTGATGTAGACCTCTGATCTCCACACTACATATACATGCCCTTAGATGAACATATA 76083

Qy       947 ------------------------------------------------------------ 947

Db     76082 CAGACATGACACACACATGTGTGCACCTAGAGCCCACCTTGGTTATACATATATACACCA 76023

Qy       947 ------------------------------------------------------------ 947

Db     76022 TGAGAGCCCTTGTTGCCAGGAGGAAAACAACTTCACTGAAGAGCTGATTCTATAGAGTCC 75963

Qy       947 ------------------------------------------------------------ 947

Db     75962 AGTTTTGTGAGGAGTGGCAGGTTCTGTACAGCAGGGAAGACCGTGGTGGACTCAGTATGA 75903

Qy       947 ------------------------------------------------------------ 947

Db     75902 GGCCCAGTAAGATCCCGTGGCTCTCTCTATTCCAATCAGAAGGGAGTGAGCTTCCAAACC 75843

Qy       947 ------------------------------------------------------------ 947

Db     75842 AGGTGTGAACCAGGTGTGGTACTGCACACCTTTGATCTCCGTCCTGGTTTATGAGCACTG 75783

Qy       948 --------------------------------------AspSerLeuAspValHisAlaV 955
                                                   ||||||||||||||||||||||
Db     75782 CTTCCTTCTGACCTGGAGTCACACTGTTCTCCCCCGCAGACTCCTTGGATGTGCACGCCG 75723

Qy       955 alThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysA 975
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75722 TCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTG 75663

Qy       975 laGlyLeuArgGlnSerSer---------------------------------------- 981
             |||||||||||:::||||||
Db     75662 CTGGCCTACGACGAAGCTCCCACAGGTGAACCGGGGAACCACTGAGACCACTGAGTTACC 75603

Qy       981 ------------------------------------------------------------ 981

Db     75602 AGGAAGGGCTGCAGAGTGGACAAAGGAGCGTGTCCTTCCCTGCTGCCCCGGAAGCCCCAG 75543

Qy       981 ------------------------------------------------------------ 981

Db     75542 TATAGATAGGGCTATGGAGCTGGAGTGAAGTTCAGTGCTACATGTGCCCCCATGTGAGGC 75483

Qy       981 ------------------------------------------------------------ 981

Db     75482 CCTGGGCGTCATCCTCAGTAGCACAAAACAAGTGGATAATTTATTTATTAGGATTCTGAG 75423

Qy       981 ------------------------------------------------------------ 981

Db     75422 ATGAAGGATTGAGGTGTTGTTCTCATGAGAATGTAGAATAAACTTAGGGAAGCCCCAGGA 75363

Qy       981 ------------------------------------------------------------ 981

Db     75362 CTCCCCACCTCAGTAACAGCGGGCTGAGCTACCTGTGGGCTTCCCTCATGGCTTAAATGG 75303

Qy       981 ------------------------------------------------------------ 981

Db     75302 AAATGGCCCATCTCGCCATACAAGGAAGCGTGCCATCTGAGGGGTCAGGCCGAGTGGACT 75243

Qy       981 ------------------------------------------------------------ 981

Db     75242 TGTATTCCTTCCTTTCTGTTTGATTGCACCACACCAGGAAACCAAATCCACTCTTGTCCG 75183

Qy       982 -----------------------HisArgGluAlaLeuProGlyLeuSerSerSerGlyT 994
                                    |||||||||||||||||||||||||||||||||||||
Db     75182 ACCCATAACTTTGTTCTGTTGATCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCA 75123

Qy       994 hrProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaA 1014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75122 CCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTG 75063

Qy      1014 laHisGluLeuGluSerLeuValHisProArgProGlnAspTrpSerProProProSerA 1034
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75062 CCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAG 75003

Qy      1034 spValGluAspLysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgG 1054
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75002 ATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGG 74943

Qy      1054 lyHisSerLysArgLys------------------------------------------- 1059
             |||||||||||||||||
Db     74942 GCCACTCCAAGAGAAAGGTGAGACTGGAGTCTCTTCAGGTCAGGGACTCACACGAACCCA 74883
```

```
Qy       1059 ------------------------------------------------------------ 1059

Db      74882 TTCCCAAATCAAACACCAGGGGGCACCAGAGCACAACTTCTGACGGTGAGATGCGCTCGC 74823

Qy       1059 ------------------------------------------------------------ 1059

Db      74822 GCGCACACACACACACACACACACACACACACACACACACACACACACAGGCACTGAACC 74763

Qy       1059 ------------------------------------------------------------ 1059

Db      74762 ATACAGAGAAAGCCTGTCCATACCACCTGGGAGGATCCCAGAGCTGTGTTTTCGAGAGCA 74703

Qy       1059 ------------------------------------------------------------ 1059

Db      74702 CCCATACTTCTCTTCACTTACCTAATATGGAGAAAAGAGGCTGGTCACTTTTGCTGTGGC 74643

Qy       1059 ------------------------------------------------------------ 1059

Db      74642 CAAGTAGGGCTGGCCCCAGAATAAGGTGAAACCCCATAGGCCAGGTGTTCTGGGGGCACA 74583

Qy       1059 ------------------------------------------------------------ 1059

Db      74582 CCCTGACCAAGTGGACAGTCAGTCTTCCCATCTCTTCACTTTCCTCTGCAATTAGAAAGA 74523

Qy       1059 ------------------------------------------------------------ 1059

Db      74522 GTCCAGCAAGCTTGGGAGCTTTGGATTTTGCACATCAACTGTTGTTAGGCTTATGGGTCC 74463

Qy       1059 ------------------------------------------------------------ 1059

Db      74462 GCCCTTCCCATTCTCCCTACCTCAGTTCTCACCTCTGTCCCGGGGTGGGGCTTTAGTGTG 74403

Qy       1059 ------------------------------------------------------------ 1059

Db      74402 GAGTTTAGTGTTATCTGGAGATAACAGGGGTCCACAGGTCTGGACCCTGCATGCCCATAC 74343

Qy       1060 ---------------------IleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGly 1072
                                   |||||||||||||||||||||||||||||||||||||||
Db      74342 TGTCTTCTTCCTTCTCTGCAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGC 74283

Qy       1073 SerTrpAlaGlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAla 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74282 TCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCT 74223

Qy       1093 LeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGly 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74222 CTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTATATGACGCC 74163

Qy       1113 IleLys------------------------------------------------------ 1114
              ||||||
Db      74162 ATAAAGGTTGGTTGGTGCTGTTTGAGAGGGAGGGAGAGGCAGGATGGGGAGGCAGGAGGG 74103

Qy       1114 ------------------------------------------------------------ 1114

Db      74102 AAAGTCCCCCCTCCATGCACTTCCAGGGACAGAAGTGTATTTGAAAGACCTTGTAGCTTC 74043

Qy       1114 ------------------------------------------------------------ 1114

Db      74042 CAAAGGGGAGTGGGTGACAGGAAGCCAGTGTGGAGCTTCTTTACCTGGTCCTAAGTGACT 73983

Qy       1114 ------------------------------------------------------------ 1114

Db      73982 TACAGTAGACTTACAGAGTATCTGAAGGTGGCCTTAGCCCCTTAGATACGGCCAGCGCCT 73923

Qy       1114 ------------------------------------------------------------ 1114

Db      73922 CCTTGTCTTCGTGGAGGTTCTGGGTCTAGTACTTTCTGGTAGCCTGAGTGAGGGCAAACG 73863

Qy       1114 ------------------------------------------------------------ 1114

Db      73862 TGTGACATTCTGCGTTTCTCATCTCTGCCAGAGCCTGAGCGTCGTCTCTTCTTAGGTGGA 73803

Qy       1114 ------------------------------------------------------------ 1114

Db      73802 ACAGTGCTTGTATCATCTAATGGGGAGTCAGCCTCTCCCCCCGCCCCCCAATATGTTGGT 73743

Qy       1114 ------------------------------------------------------------ 1114

Db      73742 CATTTTCTGAAGACCATTCTCAGTGGTGCTCACCTAGCCTTTTGTTCTCCACTCTGCCAG 73683

Qy       1115 SerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIle 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73682 AGCAACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATT 73623

Qy       1135 ValHisSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGlu 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73622 GTCCACTCCGACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAG 73563

Qy       1155 ProGluGluProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSer 1174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73562 CCAGAGGAACCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCT 73503

Qy       1175 GlnGlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyPro 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73502 CAAGGAGCAGACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCA 73443

Qy       1195 GluArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGln 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73442 GAGAGGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAG 73383

Qy       1215 ProSerThrAlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysPro 1234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73382 CCCAGCACTGCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCT 73323

Qy       1235 LeuThrValSerProSerLeuProArgAlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73322 CTAACAGTCAGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 73269

RESULT 12
AC106509/c
LOCUS       AC106509           224227 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-202I8, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC106509
VERSION     AC106509.4  GI:30580817
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 224227)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S:L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
```

```
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 224227)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:22857495.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLBA
                Center clone name: CH230-202I8
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 211258 bases at least Q40
                Consensus quality: 214044 bases at least Q30
                Consensus quality: 215992 bases at least Q20
                Estimated insert size: 220399; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   221791: contig of 221791 bp in length
            *   221792   221891: gap of unknown length
            *   221892   222909: contig of 1018 bp in length
            *   222910   223009: gap of unknown length
            *   223010   224227: contig of 1218 bp in length.
FEATURES             Location/Qualifiers
     source          1..224227
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-202I8"
     misc_feature    1..1752
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3617..3835
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH334357"
     misc_feature    220120..221791
                     /note="wgs_contig"
BASE COUNT    58284 a  52300 c  50849 g  55959 t   6835 others
ORIGIN

Alignment Scores:
Pred. No.:              2.58e-114     Length:         224227
Score:                  3064.50       Matches:        984
Percent Similarity:     21.91%        Conservative:   24
Best Local Similarity:  21.39%        Mismatches:     45
Query Match:            46.40%        Indels:         3554
DB:                     2             Gaps:           18

US-09-754-997A-2 (1-1252) x AC106509 (1-224227)

Qy        171 ProIleIleThrTrpGluLysAspGlnValThrValProGluGluProArgLeuIleThr 190
              |||::::::|||          ::::::         |||       |||   |||:::
Db     111847 CCCCTTCTGACA---------GAAAGAACCATGACGCCCCTGACCCCTGCTTTGGTC--- 111800

Qy        191 LeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArg 210
              :::||||||
Db     111799 ATTCCGAAG--------------------------------------------------- 111791

Qy        211 CysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAla 230
                                          ||||||   :::               :::
Db     111790 -----------------------CAGAGGGTGGCAAGCTGACCCTGTCCTCCTATCTTC 111755

Qy        231 LeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleValAlaAlaProGlu 250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111754 TCTACAGGGTCCTTGGAGGCTACCAGGGGCCAGGATGTGGTCATAGTGGCAGCCCCAGAG 111695

Qy        251 AsnThrThrValValSerGlyGlnAsnValValMetGluCysValAlaSerAlaAspPro 270
              ||||||||||||||||||||||||:::||||||:::|||||||||||||||||||||||||
Db     111694 AACACCACAGTGGTGTCTGGACAGAGTGTAGTGTTGGAGTGTGTGGCCTCCGCTGACCCC 111635

Qy        271 ThrProPheValSerTrpValArgGln--------------------------------- 279
              |||||||||||||||||||||||||||
Db     111634 ACCCCTTTTGTGTCCTGGGTCCGACAGGGTGAGTGCTGAGGAAAGAGGAGGAGGAGATTG 111575

Qy        279 ------------------------------------------------------------ 279

Db     111574 GGGGAGTGGGCAAGGTGGTACAGTATCTCACAGGCTCCCAGTCAGGAGCTTCTACCTAAA 111515

Qy        279 ------------------------------------------------------------ 279

Db     111514 GATGGAAGGAGAGAATTGACTCCACAAAGTTGTCCTCTGGCTTCCATGAGCACTCTCAAA 111455

Qy        279 ------------------------------------------------------------ 279
```

```
Db  111454 CAAGCGCACATGTGCACACATTCATGCAGTAATAAAGTCTAAAGAAGGTTGTGTTAAGCC 111395
Qy     279 ------------------------------------------------------------ 279
Db  111394 CAGGGATGGAGGCGCACATCCTATCCCAAGCACTCAGGAGGCAGAGGCAGGCAGATCTCT 111335
Qy     279 ------------------------------------------------------------ 279
Db  111334 GAGTTTCAGGGCAGCCGGGACTACTCAGAGAAACCCTGCCTCGAAAAATTAGAGGGCAGG 111275
Qy     279 ------------------------------------------------------------ 279
Db  111274 GGCGGGGAGAGAGAGACAGAGACAGAGACAGAGACAGAGACAGAGACAGGCAGACAGACA 111215
Qy     279 ------------------------------------------------------------ 279
Db  111214 GACACAGAGAGAGAGAGACAGAGACAGAGACAGAGACAGACACAGAGAGACAGACAGAGACA 111155
Qy     279 ------------------------------------------------------------ 279
Db  111154 GAGAGACAGAGACAGACAGAGACAGATAGAGACAGTGGGAGCAGGGAAGGAAGGGGAAAG 111095
Qy     279 ------------------------------------------------------------ 279
Db  111094 GGACAAGAGTTGTATTAAATATCTGCCCGCCCACCAGAGGCCATTTAGCCGTCGTTTACC 111035
Qy     279 ------------------------------------------------------------ 279
Db  111034 CACCTCCCATCTGCCTTTCCACCTAGGGATCAGAGTTAGAAGGAACTGGATTCAAGCCAG 110975
Qy     279 ------------------------------------------------------------ 279
Db  110974 CGACTGTGGTGCAGCTTTGCGGGGTTCGGGAGTGGCGGTGTAGTAGGATCTTTGTAGTTG 110915
Qy     279 ------------------------------------------------------------ 279
Db  110914 GAGAGGGTGTAGGGGGACAGTTCCAGCGTCCCCCTCTAACCTACTCCTTCTCTGTCACCA 110855
Qy     280 AspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAla 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110854 GATGGGAAGCCTATTTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCC 110795
Qy     300 SerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArg 319
           ||||||||||||||||||||||||||||||||||||||||||||||||     ||||||
Db  110794 AGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGGGCCAACAAGCCCCGCACCCGA 110735
Qy     320 AspPheAlaThrAlaAlaAlaGluLeuArgValLeu------------------------ 331
           ||||||||||||||||||||||||||||||||||||
Db  110734 GACTTCGCCACTGCCGCTGCTGAGCTCCGAGTGCTTGGTAAGGTCTGGCGCTGTGGGACA 110675
Qy     331 ------------------------------------------------------------ 331
Db  110674 TTAAGGACAGGTTTGTGTGTTAGAGAGAGAAGGGGCGATCCTGGAAGACTAAACTTGAGG 110615
Qy     331 ------------------------------------------------------------ 331
Db  110614 GGCACGACATGGAGGAGTCTGAATTTACTGATTGATGGTGATTGTTGGAGGACTGGATT 110555
Qy     331 ------------------------------------------------------------ 331
Db  110554 CGGTAGCTGGCCGGGGCGGGGGGGTGGGGTGGAGGTGGATGGGGCGAATTGAGAGAGGA 110495
Qy     331 ------------------------------------------------------------ 331
Db  110494 ATGGAGGAGAGGGGTATTTAAAGCCTGGCTTCTGTCATGACTAGTGAAAGCGTGGGAGAG 110435
Qy     331 ------------------------------------------------------------ 331
Db  110434 CAGGAGTGGGGTGGGGTTATGGGAGGAGATAGTGGAATGGTTTCTGAGAGGGTATGAAGC 110375
Qy     331 ------------------------------------------------------------ 331
Db  110374 TCAGGGGCCAGAATCTGTACGTGGTAGCATTTGGAAAGCTGTAGCAGTAATGTGAGTTCT 110315

Qy     331 ------------------------------------------------------------ 331
Db  110314 ATGCCAACCCGACCTACAGCTACAGAGTGAAAAGCCTATCTCAAAACAAACTTATCAAAC 110255
Qy     331 ------------------------------------------------------------ 331
Db  110254 AAACAAACAAACAAACAAATCAGCTGGAGTGGTGCTGAGGTGGACAGATCAGAAATTCAA 110195
Qy     331 ------------------------------------------------------------ 331
Db  110194 GGCCAACTTGGGCTACAAAAGAGATTTCTCAATACAACAAAAAACTGGGGGCCAGGCTTG 110135
Qy     331 ------------------------------------------------------------ 331
Db  110134 CTGAAAGCTGGAGTAACTGCTGGATATAGGAGACAATGGGGAGGAGGGAGTGAAGTTACA 110075
Qy     331 ------------------------------------------------------------ 331
Db  110074 GCCTAGAGCTTTAGAGCTGGAAATCTGCTGGGCGCTGGGAGAAAGAGGGCCTGTGGCCGA 110015
Qy     331 ------------------------------------------------------------ 331
Db  110014 ATGGGCAGGCGGGTGGACCAGAGCTGGGTCCGGGTGACTAGGAGAGAGGACCTTAGAGTG 109955
Qy     331 ------------------------------------------------------------ 331
Db  109954 AGGCTGGGGGCGCGGTGTTGGGACACTCGGGGAGGGGTAGGCCCTCTTGTCAGAGTCTCC 109895
Qy     332 -----------AlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArg 347
                       ||||||||||||||||||||||||||||||||||||||||||||||||
Db  109894 CTCCTGCCTCCAGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTGTCGCGGACCAGG 109835
Qy     348 AlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrp 367
           ||||||||||||||||||||||||||||||||||||||||||||||||||     ||||||||
Db  109834 GCCAGCACTGCGCGCTTCGTTTGCCGCGCGTCCGGGGAGCCACGACCTACGCTGCACTGG 109775
Qy     368 LeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGly 387
           ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||||
Db  109774 CTGCACGACGGGGCCCCGTTGCGACCCAACGGCGCGTCAAGGTGCAGGGCGGTGGCGGC 109715
Qy     388 SerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGlu 407
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db  109714 AGCTTGGTCATCACTCAGATCGGCCTGCAGGACTCTGGTTACTACCAGTGCGTAGCGGAG 109655
Qy     408 AsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGlyLeu 427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109654 AACAGCGCGGGAACCGCCTGTGCGGCTGCGCCCCTGGCGGTGGTAGTGCGCGAGGGGCTG 109595
Qy     428 ProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAla 447
           ||||||||||||||||||     ||||||||||||||||||||:::|||||||||||
Db  109594 CCCAGTGCCCCGACTCGGGTCAAGGCAACGCCGCTGAGCAGCTCCGCTGTGCTAGTGGCC 109535
Qy     448 TrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLys 467
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109534 TGGGAGCGGCCTGAATTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAG 109475
Qy     468 AlaArgGlyVal------------------------------------------------ 471
           ||||||||||
Db  109474 GCAAGGGGTAT-GTCTACTCTTGACCACGGGACCCCAGTTAGGGCAAGTACGGGATTGAT 109416
Qy     471 ------------------------------------------------------------ 471
Db  109415 TAGTAAAGAATTGCCTAATGGTGACCTCTGCAGATGAGGGGAAGGGATGGAGAGCAATCC 109356
Qy     471 ------------------------------------------------------------ 471
Db  109355 TTCAGGGAAGTGGTCTTCAGTCTAGGATTAGATTATAGAGTGTACACACACACACTCCTT 109296
Qy     471 ------------------------------------------------------------ 471
Db  109295 CCCGGAGAGAATGCTTTACAGGGTTAAAGGAGCTGTCGTCCACCCCAGCTAACTGCAGGA 109236
```

```
Qy        471 ------------------------------------------------------------ 471
Db     109235 GGGGACACTTAGACAGCGCAGGGAACTTCCTGGCCTATCCTGGAATTAATACAATGGATA 109176
Qy        471 ------------------------------------------------------------ 471
Db     109175 GACTCGATGGCATTTATATTCTCCACCCTAGATTGTCTGTGACAATGTAGCTATACATAG 109116
Qy        471 ------------------------------------------------------------ 471
Db     109115 ACCAGTAGAATCCCCAGTCTCCAGCTGGTCAGATAGGGAGTGGGCTGTGACAGTCCCTTT 109056
Qy        472 ---------------------------AspAsnValGluTyrGlnPheAlaValAsnAsn 482
                                         |||||||||||||||||||||||||||||||||
Db     109055 GTCCCCCGTGCTACCACCTCAGGAGTAGACAATGTGGAGTACCAGTTTGCAGTAAACAAT 108996
Qy        483 AspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrVal 502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108995 GACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTACGAATTCTACGTG 108936
Qy        503 ValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeu 522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108935 GTGGCCTACTCCCAGCTAGGGGCCAGCCGGACCTCCAGCCCAGCCCTGGTGCATACACTG 108876
Qy        523 AspAsp------------------------------------------------------ 524
              ||||||
Db     108875 GACGATGGTAGGACCTCTGAACCTGCAGTGACCTGGGACCCAGAGACACTCCCAAGCAGG 108816
Qy        524 ------------------------------------------------------------ 524
Db     108815 GACAGTGTGGCCCAAGGCCCTCTTCCCTGCCTTGTAATGTCTGTTGTGCCCTATAAAGCA 108756
Qy        524 ------------------------------------------------------------ 524
Db     108755 ATGGGCAGAAGTTCTTCTCCCAGTGTAAAAATGGGGAAACTGAGGCTCAGAAAATGAAAC 108696
Qy        524 ------------------------------------------------------------ 524
Db     108695 GACTTGCCAAGCTCAGCAGGACAGCTGGGGCTTGAGCCCAGCTCTCCAGTCCCTCTACAG 108636
Qy        524 ------------------------------------------------------------ 524
Db     108635 TGCAGCCCTTCTTCCTAATGCTCATGTACCATCCAGCCATTGACTCGAAAGCTTCTGGGC 108576
Qy        524 ------------------------------------------------------------ 524
Db     108575 CACACTGTTCACTAGCCTTGGGATAGAATGGATGTCTCCTCATCCTCTAGCCTAGGGCTC 108516
Qy        524 ------------------------------------------------------------ 524
Db     108515 CCCCGAAAGGCTCCCTTTCCTCCCTCTGATCCCGGCCCCCGGCCTACTTCTGACTTGCTG 108456
Qy        524 ------------------------------------------------------------ 524
Db     108455 TGTGACCTGGACAAGTCTCTGACCACTCCTGACCTTGTTTCCCTGTCGGTGGAGAGGAGT 108396
Qy        525 ------------------------ValProSerAlaAlaProGlnLeuThrLeuSerSer 536
                                      ||||||||||||||||||||||:::|||||||||
Db     108395 TCATGCCAATGTTCTTTTGTCTCAGTCCCCAGCGCAGCACCCCAGCTCTCCCTGTCCAGC 108336
Qy        537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108335 CCCAACCCTTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGA 108276
Qy        557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLys--------------------- 569
              ||||||:::|||||||||||||||||||||    ||||||
Db     108275 CAGGTGGTGAAGTACAAGATAGAGTACGGTTCAGGGAAGGAAGGTGAGTGGGGCTGAGGC 108216
Qy        569 ------------------------------------------------------------ 569
Db     108215 AGTGCAGAGATGACTGTGGGAATGGGGGTCCTCCCCTAGGTAGGGGGAGTTGTCAGTGGG 108156
Qy        569 ------------------------------------------------------------ 569

Db     108155 AGGGGGTCCTCCCCTAGGTGGGGGGAGTTGTCAGTGGGAATGGGGCCTAATGGGGGTATG 108096
Qy        569 ------------------------------------------------------------ 569
Db     108095 GTAAAGCTCTGGAAAAAGAGTGCCAAGAGACAGATAGATAGGTTGGTGCAAGCCCGAAAG 108036
Qy        569 ------------------------------------------------------------ 569
Db     108035 ACTGAGTTCAATCTCAGACTTCATGTTTTAAAGGTGGGGGGGAAGGCTGTCACAGTGGTA 107976
Qy        569 ------------------------------------------------------------ 569
Db     107975 GTGCCTGTTTGTAATCCCAGTGCAGGGGAGGTGGAGACAGGAGGACACCTTGAGGTCATT 107916
Qy        569 ------------------------------------------------------------ 569
Db     107915 GGGCTAAAGCCAGAGTTCCAAACCAGTGAGAGACCCTGTGCAGAAGAGAAGGTGGACAAT 107856
Qy        569 ------------------------------------------------------------ 569
Db     107855 ACCAAAGAAATGACAAGAGGTTACCATCTGACCTCTACATGTGTGCGCACACACATGTAC 107796
Qy        569 ------------------------------------------------------------ 569
Db     107795 CTGCACATATGTGAAAACTCACACTCATGGACACACACACAGAAACAGTGTTTGGGGGCA 107736
Qy        569 ------------------------------------------------------------ 569
Db     107735 GTTGAAGAAAAGGGAAGACAAAGAGGGTCAGTAGCTTGGGCATGGGTAAGAAGTATAATC 107676
Qy        569 ------------------------------------------------------------ 569
Db     107675 ACTGGAGCCTTGGACTCAGTGTCCCATCCTTTGAACCCTTCTAGGTTTTTGCTTTTGTTT 107616
Qy        569 ------------------------------------------------------------ 569
Db     107615 GTCTTTTAAGATTTATTTATTTTATGTATATGAGTGCTTTATTTGCATGCACACCTGCAG 107556
Qy        569 ------------------------------------------------------------ 569
Db     107555 GCCAGAAGAGGGCATCAGATCCCATTACAGATGGTTGTGAGCCATGGTGTGGTTTCTGGG 107496
Qy        569 ------------------------------------------------------------ 569
Db     107495 GATTGAGCTCAGGACTTTTGGAAGAGCAGCCAGTGCCCCTAACAATTGAGCCATCTCTCC 107436
Qy        569 ------------------------------------------------------------ 569
Db     107435 TGCCCCAGAAACCTTCTAGATTAAACACCAGAGAAGAGGGGGGAGGCTGCAGGAGTGCAC 107376
Qy        569 ------------------------------------------------------------ 569
Db     107375 ATGGGGCATGGGCTGGTCCTACTGGGGAGATGATGCTGGTCTGGGGAAGGCGATGTGGTT 107316
Qy        569 ------------------------------------------------------------ 569
Db     107315 TGGAAAGGAATGGAGAGTATGGAAAAGAAAAGTGAGGAGGGTCAGGGTAAAGATAAAGAG 107256
Qy        569 ------------------------------------------------------------ 569
Db     107255 AGAAGCAGGAGGCTGACCCAAGGCAGCAGTCAGGGCTCCTGACACCAGGTACTGCTTAGT 107196
Qy        569 ------------------------------------------------------------ 569
Db     107195 GCTCTTAGCTAATCATCAAACCTGTACAAGGGGTGTGGACTGGCTGCTGCTATTGGTTCA 107136
Qy        569 ------------------------------------------------------------ 569
Db     107135 TAGGGGAGGAGACCATAGCCGTGGTCTTTCCATGTTGAACCAAGGCCAATGGGTTTCAAC 107076
Qy        569 ------------------------------------------------------------ 569
```

```
Db  107075 AACTATAGATAGAAGGCTCAAAGTCAACCTGGAGGCACGAACCCTGCCTCCGTGCTGAGG 107016

Qy     569 ------------------------------------------------------------ 569

Db  107015 GGTTAGTGCCCAGAGGACCCGGGGACCTGAAATGGAAGCATCAGTGTCCTTTTAGAGACC 106956

Qy     570 ----------------------------------------------GluAspGlnValPh 574
                                                         ||||||||||||||
Db  106955 AATGGTGTCCCAGTGAAGACCCCAGCAGCCTATCCTTTTTGTCCCAGAAGATCAGGTTTT 106896

Qy     574 eSerThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLy 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  106895 CTCCACTGAGGTGCCGGGAAATGAGACTCAACTTACGTTAAACTCACTCCAGCCAAACAA 106836

Qy     594 sValTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTr 614
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  106835 GGTATACCGAGTCCGGATTTCAGCTGGCACTGGGGCTGGCTACGGAGTCCCTTCTCAGTG 106776

Qy     614 pMetGlnHisArgThrProGlyValHisAsnGlnSerHisVal----------------- 628
           |||||||||||||||||||||||||     |||||||||||||
Db  106775 GATGCAGCACAGGACACCTGGTGCACACAACCAAAGCCATGG-TATGGAGCTGCGTGGGC 106717

Qy     628 ------------------------------------------------------------ 628

Db  106716 AGACTGGACAGTGGGCACAGGTGCTCTATGGCCTTGGAACCAGGGCCACCCCCCTGCACA 106657

Qy     628 ------------------------------------------------------------ 628

Db  106656 GACCTGTTGTTGCCCCCACACATACACAAGCACTTGATGCACCAGGGAGGCCTGGGTGTA 106597

Qy     628 ------------------------------------------------------------ 628

Db  106596 GGAGTAAAGGTCTGGTCTCAAATATCTCCCTATTCACTCTGTGACTTTGACAAGCCACTC 106537

Qy     628 ------------------------------------------------------------ 628

Db  106536 ACATCTCTAAAGCTTCGTTCCTACCAAGGGGTGGCAGCAGGATTAGCGATGTAGTTATAC 106477

Qy     628 ------------------------------------------------------------ 628

Db  106476 ACAAAGTGGTTGTTGGCACAGAATAGGAGCTCCTTGGGCTGAGACCATGGCTCACTTGGC 106417

Qy     628 ------------------------------------------------------------ 628

Db  106416 AGAGTGCTTTCCTAGCATGCACAGAGCCCTAGGTTCAAGCCCCATGAACCCCAGCATGCC 106357

Qy     628 ------------------------------------------------------------ 628

Db  106356 AGAAATTCAAGGTCTGTCTCACCTACATAGTGAGTTTGGGACCAGCCTAGGAGGCATGAA 106297

Qy     628 ------------------------------------------------------------ 628

Db  106296 ACTCTGTTTCAGAGAAAAGAAGAAACCCATTTAAAAAAAAGTTGCTGGCTTTGCCAAGAG 106237

Qy     628 ------------------------------------------------------------ 628

Db  106236 GCTGAATAGCACACTGGTTAAATGGCCACAAAGTCTCCAGGAGAGTGGTAGTGGAGGCGC 106177

Qy     628 ------------------------------------------------------------ 628

Db  106176 ATGCCTTTAATCCTAGCATTCCAGAGGCAGAGGCAGGCAGATCTCTGAATTCAAAGACAG 106117

Qy     628 ------------------------------------------------------------ 628

Db  106116 CCTCGTCTACAGAGCAAGTTCCAGGAGAGCCAAGGTTACATGGAGAAACCCTGGCTCGAA 106057

Qy     628 ------------------------------------------------------------ 628

Db  106056 AAAGAAAAAGAGAAAAAAAAGAGAAGATTATAAAGTTGGCTACCATATTCAGACATTAGA 105997

Qy     628 ------------------------------------------------------------ 628

Db  105996 CAATCTATAGCTTTTAAAAATTATAGACTTATAAGGCTGCCAAGATGGCCCAGCTAACAA 105937

Qy     628 ------------------------------------------------------------ 628

Db  105936 AGGGGCTGGCTCCCTTTTCTGGGGATCTGACTCCTATTGTCTCCCCCAGAATGTATACGG 105877

Qy     628 ------------------------------------------------------------ 628

Db  105876 TGAAAGAAATAAGTCCACACCCTGCGGACTCTGACCTCTGACCGCCACATGCACACACAT 105817

Qy     628 ------------------------------------------------------------ 628

Db  105816 AAATTAAGTAATTTAAAAAAATTTAAATACATATACAAAATGTTTGAAACAGAGTCCAGC 105757

Qy     628 ------------------------------------------------------------ 628

Db  105756 ATACAGCCAACTCTCTACTCTTAGGCTCAGAGGATTGCTCATCACGCTAATGGTTCCAAA 105697

Qy     628 ------------------------------------------------------------ 628

Db  105696 CCTCCCTCATGATCTCAACAGCCCCTTACCTGTCCTTCAAGAGACCTACAGGGGCTGGGG 105637

Qy     628 ------------------------------------------------------------ 628

Db  105636 ATTTAGCTCAGTAGTAGAGCGCTCACCTAGGAAGCGCGAGGCCCTGGGTTCGGTCCCCAG 105577

Qy     628 ------------------------------------------------------------ 628

Db  105576 CTCCGGGAAAAAAAAAAAAAAAAAAAAGAGAGACCTACAGAAACCCCTCTGTTCCCTAACC 105517

Qy     628 ------------------------------------------------------------ 628

Db  105516 GCGGGCCTGCGCAGAGGCGCCCCCTGCTGTCTCTTCTGCATGCCCGGCAGACGCGCCTGC 105457

Qy     628 ------------------------------------------------------------ 628

Db  105456 GCAGCTAACGTGCAGAGCTCAAGTCCACCGGGAGGTGACAGAGTGGAGCAGTAGGGAAAG 105397

Qy     628 ------------------------------------------------------------ 628

Db  105396 CCCAGGCGTGACGACTCCACCACTTCACTCCTCATTCTCTTCCCTTTGTCCCCATCATCT 105337

Qy     629 ----------------ProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSe 643
                           ||||||||||||||||||||||||||||||||||||||||||||
Db  105336 TCCTCTGCCCCCAGTCCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTC 105277

Qy     643 rLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTy 663
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105276 CCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCACCCAGATCTCTGGATATAAACTCTA 105217

Qy     663 rTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGl 683
           ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105216 CTGGCGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGCCGTGG 105157

Qy     683 yAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLe 703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105156 AGATCAGGCTTGGGACGTCGGGCCGGTGCGGCTGAAGAAGAAAGTCAAGCAGTATGAGCT 105097

Qy     703 uThrGlnLeu-------------------------------------------------- 706
           |||||||
Db  105096 GACCCAGTT-AGGTGAGTAGGCCTGGGCTTGGGAGGGTCATGGGCAAGCTAAAAACAGGG 105038

Qy     707 -------------------------ValProGlyArgProTyrGluValLysLeuValAl 718
                                    ||||||||||||    ||||||||||||||||||
Db  105037 AGAGTGACCTCCATGGCCTTCCCCAGTCCCTGGCAGGCTGTACGAGGTGAAGCTAGTAGC 104978

Qy     718 aPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaPr 738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104977 TTTCAACAAACACGAGGACGGCTACGCTGCTGTATGGAAGGGCAAGACGGAGAAGGCGCC 104918

Qy     738 oThrPro----------------------------------------------------- 740
           |||||||
Db  104917 CACACCAGGTGAGGGGGTGGGGCAGTAGGGGTCTCAGAGTCTCCTTTTCTCTTTTCTTTA 104858
```

```
Qy        740 ------------------------------------------------------------ 740
Db     104857 CCAGCCAGCAGACTGTGTGCATGCCACAGCCACCCAGCACACCCTCCTGGGCAGGAGTTT 104798
Qy        740 ------------------------------------------------------------ 740
Db     104797 TCCTGTCCTAGACATGCCCCTCATTCTCAAGCTGAGCAGAGCAAATAGGGAAAGTTTGGA 104738
Qy        740 ------------------------------------------------------------ 740
Db     104737 AAGAGAAACTCCATGTCTGATGCGTGATTCCTGGCTCTAATTTTAATCTTTAGATAGTTT 104678
Qy        740 ------------------------------------------------------------ 740
Db     104677 TTAAGATCACATCCACTTATACCACACGATTCATCAGATTGTCCCATAGTTTTCTTTTAA 104618
Qy        740 ------------------------------------------------------------ 740
Db     104617 GATTTATTTATTTATTTATTTTATATACGAGAGTACACTGTACTTGTCTTTGGACACACC 104558
Qy        740 ------------------------------------------------------------ 740
Db     104557 AGAAAAGGGCATAGGATCCCACTACAGATGGTTGTGAGCCACCATGTGTTTGCTGGGATT 104498
Qy        740 ------------------------------------------------------------ 740
Db     104497 TGAACTCAGCTCTCAACCACTGAGTCATCTCTCCAGCCCCATCCCATAATTCTTAAGTAC 104438
Qy        740 ------------------------------------------------------------ 740
Db     104437 AACACAGTGGTTTGTCACCTTGACAATCCATTCTGACAGTCAGAAGCCAAAACCCTGGGT 104378
Qy        740 ------------------------------------------------------------ 740
Db     104377 TCTAATTTCAGCTAGGTGATATTCTCACTCAGCAAGTGTTTTACTAAGTATCTACCACGT 104318
Qy        740 ------------------------------------------------------------ 740
Db     104317 GCCAGCTCTGTTCTGGAAAATTAGCAATGAGCAGCATGAATAAGATATCTGTTCTCTGGG 104258
Qy        740 ------------------------------------------------------------ 740
Db     104257 AATTTGTTTCATGCTTATCTAATAAACATTTTTCATCACCAAATAGTATCCCAGACAGTT 104198
Qy        740 ------------------------------------------------------------ 740
Db     104197 TTCTACCCACGCTGTAAGTATTAACTAATCCAACTTCCTAAGAACCCTCAGGAGGGGACC 104138
Qy        740 ------------------------------------------------------------ 740
Db     104137 AGTGAGATGGTTCAGCAAGTGAAGGTGCTTGGTACAAGCCAGATGACCTGGGCTTGTTTC 104078
Qy        740 ------------------------------------------------------------ 740
Db     104077 CCGGCACTCCCCAAATTGTCCTCTGACCCCCACATGCATATAGTAGCGCTCATTCCCTTC 104018
Qy        740 ------------------------------------------------------------ 740
Db     104017 CCTCCACACAACAAATACATGTGTGATCGAGAAGACGCAGTGATTGAGAATTTTAAAGAA 103958
Qy        740 ------------------------------------------------------------ 740
Db     103957 ACAAATGTAAAAAAAAAAGGAGAAGAAGAAGGAGAAGAAGAAGAAGAAGCAGCTCGAGAT 103898
Qy        740 ------------------------------------------------------------ 740
Db     103897 GTTGGTGCTGTCGTCAGATCCATATAAGAATAGAGAACAGAGCTACAGAGAAATCCCACA 103838
Qy        740 ------------------------------------------------------------ 740
Db     103837 ACTTCTCCAAAGCCACATAGTCAGTAAGTGGCAGAGAAGGAGTTTCAACCAAAGGAGCCT 103778
Qy        740 ------------------------------------------------------------ 740

Db     103777 CGTTCCAGAGTCATGGCATGCTTTCTCTGTCACATTTTCATCAGTGATGGGATGTGTCAC 103718
Qy        740 ------------------------------------------------------------ 740
Db     103717 AGTCACTCTATGTGTGACCTTAGCAAAAATTCACTTCAACTCCTGGGAGCCTCAGTTGCA 103658
Qy        740 ------------------------------------------------------------ 740
Db     103657 GCAGGAATAACATCACCCAACTTACAGGATCAAAAGAGAGGCCCGGTGCCAGCTATGGTA 103598
Qy        740 ------------------------------------------------------------ 740
Db     103597 TCCCACAGCTGTAACCCCTGTGCTAGAGGATCAGAGGCAAGAGCTAGCCTGGGCTACTAC 103538
Qy        741 -----------------------------------------------AspLeuProIleG 745
                                                             |||||||||||||
Db     103537 ATAATGAACATCTGCCCACTCTCCTTACACTTGGATTTTCCCCTCTAGATCTGCCTATCC 103478
Qy        745 lnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSerSerThrSerI 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103477 AGAGGGGGCCACCGCTGCCTCCGGCCCATGTCCACGCGGAATCAAACAGCTCCACCTCCA 103418
Qy        765 leTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsnTyrThrValA 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103417 TTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACAGTCAAGATTGTCAACTACACTGTGC 103358
Qy        785 rgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr---------- 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103357 GCTTCGGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTACTATACCAGGTATCAGC 103298
Qy        801 ------------------------------------------------------------ 801
Db     103297 AAGGGGAGGGCAACTGGAATAGTAAACTAGCGGGCTCAGCTATGGCTAGGGCACACCCAG 103238
Qy        801 ------------------------------------------------------------ 801
Db     103237 AATCTACACAGATGTCTCTTTAGAACTGCTGACCAGGGAGTAACAATGAGGGAGACTTGT 103178
Qy        801 ------------------------------------------------------------ 801
Db     103177 CTGGTCCCACCCCTAATGACCGAGTGTGTGGTCACCCTTTGGTGTGAGTGGGATGCCATC 103118
Qy        801 ------------------------------------------------------------ 801
Db     103117 CATTGGCTGGGACCCTGCCCTCCCATCATGGAGAAGATAGCTTAAGGAACTGCTGGTCTG 103058
Qy        801 ------------------------------------------------------------ 801
Db     103057 ATTGAGAAGACGCCCTGTGAAAGCCAACAGAGAGATAAACTAAGCCAAGGATAGGTAGAT 102998
Qy        801 ------------------------------------------------------------ 801
Db     102997 ACCTCAGAAGTGACTGGTCCCCCCACCCCCGGTAAAGAATAACTTAGATTAATTCAGGAA 102938
Qy        801 ------------------------------------------------------------ 801
Db     102937 GGCATCCCCGAGGAGAGAAGGGTCTGCTTTGGAGATTAGACATGGAATACCATATGTCCG 102878
Qy        801 ------------------------------------------------------------ 801
Db     102877 ACCTCTCTGTCGTTTGCCTGGCTGTGTGTGGGAGGGGAGGTTAGTCTCCCTGAGTAGCAG 102818
Qy        801 ------------------------------------------------------------ 801
Db     102817 CACCCTTCAGGTTGGGGCCTGGCTGGAGGCAGGGGCTTATGTCTGGGGTTCTGCCTGGCT 102758
Qy        802 ------------------------------------------SerSerGlyGluAspIle 807
                                                        ||||||||||||||||||
Db     102757 TGAGGCCTGTCTGAGGGCACTAGTCCTATCCTTTCTCCTGGAAGCTCTGGAGAAGACATC 102698
Qy        808 LeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyVal 827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   102697 CTCATTGGCGGGCTGAAACCATTTACCAAGTATGAGTTTGCAGTACAGTCCCATGGTGTG 102638

Qy      828 AspMetAspGlyProPheGlySerValValGluArgSerThrLeuProAspArg------ 845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102637 GATATGGATGGGCCTTTTGGCTCTGTCGTAGAACGCTCCACCCTGCCTGACCGTGAGTAT 102578

Qy      845 ------------------------------------------------------------ 845

Db   102577 CCACTGCCCTCTTCCCGCCTGTGTCACCATTGTCTGGGTCCGATGAAAGAGGAAGAAGAG 102518

Qy      845 ------------------------------------------------------------ 845

Db   102517 GAGGCTGTGCCCTCAGCCCCTAACTGTTAGTAGGGGCTGTGCCTCTGACTCTTCACCCCC 102458

Qy      846 ------------------ProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
                              |||   ||||||||||||||||||||||||||||||||||||
Db   102457 TCCCCTGCGCTGCAGGGCCCTCA-ACACCTCCTTCTGACCTGCGTCTGAGCCCCCTGACA 102399

Qy      860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102398 CCATCCACTGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAACGGCGAGATCGTGGAG 102339

Qy      880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102338 TATCTAATTCTCTACAGCAACAACCACACCCAGCCGGAGCACCAGTGGACACTGCTCACC 102279

Qy      900 ThrGlu------------------------------------------------------ 901
            ||||||
Db   102278 ACAGAGGGTAAGGACGCTCATCCCCGCCTCCTTCCCTCTTGTTCTGAGGGTTCCCGAGAG 102219

Qy      901 ------------------------------------------------------------ 901

Db   102218 CAGAGGCTGCAGCTACTGTCTGGTGCAGAGTTGGCTTGCCCATGGCGCTCTGGCCTCCTG 102159

Qy      902 ----------------------------GlyAsnIlePheSerAlaGluValHisGlyLe 912
                                        ||||||||||||||||||||||||||||||||
Db   102158 CCTACGTTCTACCTGCCACCCTACCACAGGAAACATCTTCAGTGCAGAGGTACATGGCCT 102099

Qy      912 uGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyPr 932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102098 AGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGGACAGAGGTGGGGCCTGGGCC 102039

Qy      932 oPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSer-------------- 947
            ||||||||||||||||||||||||||||||||||||||||||||||
Db   102038 CTTTTCTCGATTGCAAGATGTGATCACTCTACAGGAGACATTCTC-AGGTACCAGGAAGG 101980

Qy      947 ------------------------------------------------------------ 947

Db   101979 AGGGAGGAGGGCCTTGGGGCAGCCTGAGGTCCCAAGAGCCCTGGATTGGCTCCCAAGGCC 101920

Qy      947 ------------------------------------------------------------ 947

Db   101919 CTGCTTTCTGGAAATTGCATTTCTTTTACTTTGGAGCCCTTCGTAGTGGGGATGGGGGAG 101860

Qy      947 ------------------------------------------------------------ 947

Db   101859 CAGTGACAGCTTGATAGAGGGACTGTCAGCTTTGAGGGGTCTTTTATTATTTCATTTTCT 101800

Qy      947 ------------------------------------------------------------ 947

Db   101799 CTCTCTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 101740

Qy      947 ------------------------------------------------------------ 947

Db   101739 GTTTGGCACAGTTCATATGTGAACATCAGAGACATCAGAGAGTTGGTTTCATATGGATCT 101680

Qy      947 ------------------------------------------------------------ 947

Db   101679 CGCTTTGCCTGGAGCATTTGTGAACTCCTGAGTTCCGTGAAGAAGTGGACAGCAGGCTGG 101620

Qy      947 ------------------------------------------------------------ 947

Db   101619 AGAGACACTCACAGGGTTCAATTAAGATACTAAATGAGGGCTGGAGAGATGGCTTAGCAG 101560

Qy      947 ------------------------------------------------------------ 947

Db   101559 TTAAGAGCACTGACTGCTCTTCCAGAGGTCCTGAGTTCAATTCCCAGCAACCACATGGTG 101500

Qy      947 ------------------------------------------------------------ 947

Db   101499 GCTCACAACGATCTGTAATGGGATTGATGCCCTCTTCTGGAATGTCTGAAGACAGCAACA 101440

Qy      947 ------------------------------------------------------------ 947

Db   101439 GTGTGCTCATATATATATAAAATAAATAATTAAATAAATCTTTTTTAAAAAAGATACAA 101380

Qy      947 ------------------------------------------------------------ 947

Db   101379 AATGAGAACCCAGGGCCCCTTATTGAAAAATTATTGAAGATATGAAGATGGGCCAACGGG 101320

Qy      947 ------------------------------------------------------------ 947

Db   101319 GCACCAACCACACATAGGATCATGGAGCATGAGGGCCTGAGTTCAAATGCCCAGCATCCA 101260

Qy      947 ------------------------------------------------------------ 947

Db   101259 TGTAAAAGTTGGATGTGGCTGTGTATGCTGTAATCCTTGCTGAGGAAGCAAAGACAGGAT 101200

Qy      947 ------------------------------------------------------------ 947

Db   101199 ATCTCTGGGGCTGAAATAGTCACTTGAGGTGGACCTCTGACCTCCACACCACACATACAT 101140

Qy      947 ------------------------------------------------------------ 947

Db   101139 TCCCTTGTGTGCCTCCCCACATGAACATATGCAGACATCACACACACACGTGTAACTACA 101080

Qy      947 ------------------------------------------------------------ 947

Db   101079 CACCACACCTAGAGCCCATCTAGATTAAACATATATATGCCACGAAGCCCTTGTTGCCAA 101020

Qy      947 ------------------------------------------------------------ 947

Db   101019 GTGGAAGTCTTCTTCACCGAAAATCTGATCCTACAGAAGTCCAGTTTTGTGAAGAGAGGC 100960

Qy      947 ------------------------------------------------------------ 947

Db   100959 AGGTGGTGCATAGCAGAGAAGACCATGGTGGACTCAATATGAGGCCTACTATAGTAAGGT 100900

Qy      947 ------------------------------------------------------------ 947

Db   100899 CCTGTGGCTTTCTGTACTCCAGTCAGAAGGGAGTGAGCTTCCAAGCCAGGTGTGGTGGCA 100840

Qy      947 ------------------------------------------------------------ 947

Db   100839 CACAACTTCGATCCCAGCACTCGGGAGGCAGAGGCAGACAAATCTCTGAGTTCAAAGCCA 100780

Qy      947 ------------------------------------------------------------ 947

Db   100779 GCCTGGTCTACAGAGTGAGTTCCAGGACAGCCAGAGCTATAAAAGAATCCCTGTCTCAAA 100720

Qy      947 ------------------------------------------------------------ 947

Db   100719 ACAAACTAACTAACAAACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 100660

Qy      947 ------------------------------------------------------------ 947

Db   100659 GAGAGAGAGAAGTGTGAGCACCCCGAAGAGGAGACTAAGGCTAGCAGTGTTCCATAGAAG 100600

Qy      947 ------------------------------------------------------------ 947

Db   100599 CTCTCACCTCTGGTTTATGAGCACTGTGCCGTACTGTCTTTGAATTACACTGTTCTCCTC 100540

Qy      948 -------AspSerLeuAspValHisAlaValThrGlyIleIleValGlyValCysLeuGl 965
                   ||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db   100539 CCCACCAGACTCCTTGGATGTGCACTCTGTCACAGGCATCATCGTGGGCGTCTGCCTGGG 100480
```

```
Qy      965 yLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer----------- 981
            ||||||||||||||||||||||||||||||||||||||||:::||||||
Db   100479 CCTTCTCTGTCTCCTGGCCTGCATGTGTGCTGGCCTGCGACGAAGCTCGCACAGGTGAAC 100420

Qy      981 ------------------------------------------------------------ 981

Db   100419 GAGGGAACCATTGAGACTAAGAGAGTTACCAGGAAGGGCTGCCTCTGAAGCCCCAATATA 100360

Qy      981 ------------------------------------------------------------ 981

Db   100359 GTTAGAGCTATGGAAATGGAGTGGAGTTCAGTGCTACTTTTGCCCTCATGTGAGGCCCTG 100300

Qy      981 ------------------------------------------------------------ 981

Db   100299 GGCTCCATCATCAGTAGCACAAAAACAAGTCGATAATTTATTTATTAGCATTTTGAGATG 100240

Qy      981 ------------------------------------------------------------ 981

Db   100239 AAGGATTGGGGGCGTTGATCACATGAGACTGTAGAATAAACCAAGGGGAGCCTACCCCCC 100180

Qy      981 ------------------------------------------------------------ 981

Db   100179 CCTACCCGTCCCCCTCACCTCACAGCAGGCTGAGCTACCTGTGGGCTTTTCTCATAGCTT 100120

Qy      981 ------------------------------------------------------------ 981

Db   100119 CCCTGCTGGAAGCTTAGATGGAAAAGCTCCACGTCGCCATTCAAGGAAGGGTGCCATTTG 100060

Qy      981 ------------------------------------------------------------ 981

Db   100059 AGGGGATGGACTGTATTCCCTCCTTTCCGGTTGATCACACCACACCAGGAAGCCAAATCC 100000

Qy      982 --------------------------------------HisArgGluAlaLeuProGlyL 989
                                                  ||||||||||||||||||||||
Db    99999 TCCACTCTTGTCTGACCCATAGCTTTGTTCCTGTTGCTCACAGGGAAGCCCTCCCCGGAC 99940

Qy      989 euSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProP 1009
            ||||||||:::    |||  |||||||||||||||||:::||||||||||||||||||||
Db    99939 TGTCCTCCACGGCCACCGCTGGGAACCCAGCACTCTACTCAAGAGCTCGGCTCGGGCCCC 99880

Qy     1009 roSerValProAlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAspTrpS 1029
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    99879 CCAGTGTCCCCGCTGCCCATGAGTTGGAGTCCCTTGTGCATCCTCGTCCCCAGGACTGGT 99820

Qy     1029 erProProProSerAspValGluAspLysAlaGluValHisSerLeuMetGlyGlySerV 1049
            ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db    99819 CCCCACCGCCCTCGGATGTGGAAGACAAGGCTAAAGTGCACAGCCTTATGGGTGGCAGCG 99760

Qy     1049 alSerAspCysArgGlyHisSerLysArgLysIle------------------------- 1060
            ||||||||||||||||||||||||||||||||
Db    99759 TTTCAGATTGCCGAGGCCACTCCAAGAGAAAGGT-GAGAGGTGAGACTGGAGTCTCGGTA 99701

Qy     1060 ------------------------------------------------------------ 1060

Db    99700 GGTCAGAGGCTCACACGGACCCATTCCCAAATCAAACACTAGGGGGCGCCAGAGCACAAC 99641

Qy     1060 ------------------------------------------------------------ 1060

Db    99640 TTCTGACGGCGAGACGCACACCAGGCCCTGAGCTATCTAGAGAAAGTCTGTCTGCCCCAC 99581

Qy     1060 ------------------------------------------------------------ 1060

Db    99580 CTGGGAGGGCACCAGAGCTGCGAGCACCCATACTTCTCCTCACTAACCTAATATGGAGAA 99521

Qy.    1060 ------------------------------------------------------------ 1060

Db    99520 AAGAAGCTGGTCCCTTTGCTGTGGCCAAATAGGGCTGGCCCCAGGATGAGGTGAGATACC 99461

Qy     1060 ------------------------------------------------------------ 1060

Db    99460 GTAGGCCAGGTTCCCTGGTGGCACACCCTGACCAAATGGACAGTCAGTCCTCCCATCTCT 99401

Qy     1060 ------------------------------------------------------------ 1060

Db    99400 CCACTTTCCTCTGCAATAAGGAAGAGTCCACCAGGTTTGGGAGCTTCCGATTTTGCGATT 99341

Qy     1060 ------------------------------------------------------------ 1060

Db    99340 TTGCACATCAAGTGTTATTAGGCTTATGACTCCGCCCTTCCTATCCTCCCAACCTCAGTT 99281

Qy     1060 ------------------------------------------------------------ 1060

Db    99280 CTTACCCCTGCCTGGAGGAGGGGCTTTAGCGCAGAACCTATCTGGAGATAACAGGGGCCC 99221

Qy     1061 --------------------------------------------------SerTrpAlaG 1064
                                                              ||||||||||
Db    99220 ACAGGTCTGGACCCTGCATACCATACTATCTTCTTCTTTCTCTGCAGATCTCCTGGGCTC 99161

Qy     1064 lnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerG 1084
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    99160 AGGCAGGGGGACCAAATTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGGAGTG 99101

Qy     1084 lyProArgProAlaLeuThrArgAlaLeuProProAlaGlyThrGlyGlnThrLeuL 1104
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    99100 GTCCAAGGCCGGCTCTAACCCGGGCTCTGCTGCCTCCAGCCGGAACTGGGCAGACACTGC 99041

Qy     1104 euLeuGlnAlaLeuValTyrAspGlyIleLys---------------------------- 1114
            |||||||||||||||||||||       |||
Db    99040 TGCTGCAGGCTCTGGTGTATGATGCCAT-AAAGGTTGGTCAGTGTTGTGAGAGGAAGAGG 98982

Qy     1114 ------------------------------------------------------------ 1114

Db    98981 GAGGCAGGAGGGGAAATCCCTATGCATTTCCAGGGACAGAAAGGTATTTGGAAGACCTTG 98922

Qy     1114 ------------------------------------------------------------ 1114

Db    98921 TAGCTTCCAAAGGGGAGTGGGTGACAGGAAGCCAGTTTGGAGCAGTCTCTTCTTTACCTG 98862

Qy     1114 ------------------------------------------------------------ 1114

Db    98861 GACCTAATGGCTCATAGTAGACTTACAGAGTATCTGAAGGTGGCCCTGACCCCTCAGATA 98802

Qy     1114 ------------------------------------------------------------ 1114

Db    98801 CGGTCAGCCCCTGCTGCTCTTTATGGAGGTCCTGTGCTTAGTACTTTCTAGTAGCCTGAG 98742

Qy     1114 ------------------------------------------------------------ 1114

Db    98741 TGAGGGCGAACGCGGCATCCTGTACTTCTTGCCTCTGCAAGAGCCTGAGCATCTCAGGCG 98682

Qy     1114 ------------------------------------------------------------ 1114

Db    98681 GAACTGTGCTTGTGTTGTCCAATGGGGAGTCAGCCTCTCCCCACCCTACATGTGTTTGAT 98622

Qy     1115 --------------------------------------------------------SerA 1116
                                                                    ||||
Db    98621 CATTTCTCTGAAGACCATCCTCAGAGACGCTCACCTTCTGTTCCCCACTCTGCCAGAGCA 98562

Qy     1116 snGlyArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValH 1136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    98561 ACGGGAGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAGGCCGAAGTCATTGTCC 98502

Qy     1136 isSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProG 1156
            ||||||||||   |||||||||||||||||||||||||||||||||||||||||||||||
Db    98501 ACTCCGATTTCAGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCCG 98442

Qy     1156 luGluProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnG 1176
            ||||  |||  ||||||  ||||||  |||:::|||  ||||||||||||||||||||||
Db    98441 AGGAGGCCCTGCCTGCAGAGGTTCTGCCTCCCACTGCTGGGGTCGTGGATTTGTCCCAAG 98382

Qy     1176 lyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluA 1196
            |||||||||||||||||||||||||||||  ||||||||||  ||||:::||||||||||
Db    98381 GAGCAGACTGGCTGGGCAGGGAGCTGGAAGAGTGCCAACCAGCAACCACTGGGCCAGAGA 98322

Qy     1196 rgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProS 1216
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      98321 GGCTCACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCA 98262

Qy       1216 erThrAlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuT 1236
                ||:::|||:::|||||||||||||||||   |||||||||||||||||||||||
Db      98261 GCAGTGCTGTAGAGGAGGCCCCTGGGAAAGGCTGCCAGCCCAAAGCCCTGTGTCCTCTAG 98202

Qy       1236 hrValSerProSerLeuProArgAlaProValSerSerAlaGlnValPro 1252
                 :::||||||||||||||||||||||||||||||||||||||||||||
Db      98201 AAATCAGCCCAAGCCTTCCCAGAGCCCCTGTCTCCTCTGCTCAGGTGCCC 98152

RESULT 13
AC068507/c
LOCUS       AC068507               176282 bp    DNA     linear   HTG 06-MAY-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-349H15 map 15, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC068507
VERSION     AC068507.2  GI:8389522
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349H15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 9, 2000 this sequence version replaced gi:7684421.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10275
                Center clone name: 349_H_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156826 bases at least Q40
                Consensus quality: 166608 bases at least Q30
                Consensus quality: 170674 bases at least Q20
                Insert size: 173682; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1       1477: contig of 1477 bp in length
            *     1478       1577: gap of 100 bp
            *     1578       2785: contig of 1208 bp in length
            *     2786       2885: gap of 100 bp
            *     2886       4859: contig of 1974 bp in length
            *     4860       4959: gap of 100 bp
            *     4960       6416: contig of 1457 bp in length
            *     6417       6516: gap of 100 bp
            *     6517       8739: contig of 2223 bp in length
            *     8740       8839: gap of 100 bp
            *     8840      10868: contig of 2029 bp in length
            *    10869      10968: gap of 100 bp
            *    10969      13737: contig of 2769 bp in length
            *    13738      13837: gap of 100 bp
            *    13838      15329: contig of 1492 bp in length
            *    15330      15429: gap of 100 bp
            *    15430      17938: contig of 2509 bp in length
            *    17939      18038: gap of 100 bp
            *    18039      19981: contig of 1943 bp in length
            *    19982      20081: gap of 100 bp
            *    20082      23587: contig of 3506 bp in length
            *    23588      23687: gap of 100 bp
            *    23688      27447: contig of 3760 bp in length
            *    27448      27547: gap of 100 bp
            *    27548      30161: contig of 2614 bp in length
            *    30162      30261: gap of 100 bp
            *    30262      34737: contig of 4476 bp in length
            *    34738      34837: gap of 100 bp
            *    34838      40101: contig of 5264 bp in length
            *    40102      40201: gap of 100 bp
            *    40202      44039: contig of 3838 bp in length
            *    44040      44139: gap of 100 bp
            *    44140      49950: contig of 5811 bp in length
            *    49951      50050: gap of 100 bp
            *    50051      55555: contig of 5505 bp in length
            *    55556      55655: gap of 100 bp
            *    55656      60496: contig of 4841 bp in length
            *    60497      60596: gap of 100 bp
            *    60597      68544: contig of 7948 bp in length
            *    68545      68644: gap of 100 bp
            *    68645      76940: contig of 8296 bp in length
            *    76941      77040: gap of 100 bp
            *    77041      85480: contig of 8440 bp in length
            *    85481      85580: gap of 100 bp
            *    85581      93905: contig of 8325 bp in length
            *    93906      94005: gap of 100 bp
            *    94006     108063: contig of 14058 bp in length
            *   108064     108163: gap of 100 bp
            *   108164     123496: contig of 15333 bp in length
            *   123497     123596: gap of 100 bp
            *   123597     139965: contig of 16369 bp in length
            *   139966     140065: gap of 100 bp
            *   140066     176282: contig of 36217 bp in length.
FEATURES             Location/Qualifiers
     source          1..176282
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349H15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1477
                     /note="assembly_fragment"
     misc_feature    1578..2785
```

```
                          /note="assembly_fragment"
     misc_feature         2886. .4859
                          /note="assembly_fragment"
     misc_feature         4960. .6416
                          /note="assembly_fragment"
     misc_feature         6517. .8739
                          /note="assembly_fragment"
     misc_feature         8840. .10868
                          /note="assembly_fragment"
     misc_feature         10969. .13737
                          /note="assembly_fragment"
     misc_feature         13838. .15329
                          /note="assembly_fragment"
     misc_feature         15430. .17938
                          /note="assembly_fragment"
     misc_feature         18039. .19981
                          /note="assembly_fragment"
     misc_feature         20082. .23587
                          /note="assembly_fragment"
     misc_feature         23688. .27447
                          /note="assembly_fragment"
     misc_feature         27548. .30161
                          /note="assembly_fragment
                          clone_end:T7
                          vector_side:left"
     misc_feature         30262. .34737
                          /note="assembly_fragment"
     misc_feature         34838. .40101
                          /note="assembly_fragment"
     misc_feature         40202. .44039
                          /note="assembly_fragment"
     misc_feature         44140. .49950
                          /note="assembly_fragment"
     misc_feature         50051. .55555
                          /note="assembly_fragment"
     misc_feature         55656. .60496
                          /note="assembly_fragment
                          clone_end:SP6
                          vector_side:right"
     misc_feature         60597. .68544
                          /note="assembly_fragment"
     misc_feature         68645. .76940
                          /note="assembly_fragment"
     misc_feature         77041. .85480
                          /note="assembly_fragment"
     misc_feature         85581. .93905
                          /note="assembly_fragment"
     misc_feature         94006. .108063
                          /note="assembly_fragment"
     misc_feature         108164. .123496
                          /note="assembly_fragment"
     misc_feature         123597. .139965
                          /note="assembly_fragment"
     misc_feature         140066. .176282
                          /note="assembly_fragment"
BASE COUNT     44272 a   43276 c   42765 g   43361 t    2608 others
ORIGIN

Alignment Scores:
Pred. No.:              9.29e-94      Length:         176282
Score:                  2557.00       Matches:        877
Percent Similarity:     20.85%        Conservative:   45
Best Local Similarity:  19.83%        Mismatches:     111
Query Match:            38.72%        Indels:         3397
DB:                     2             Gaps:           18

US-09-754-997A-2 (1-1252) x AC068507 (1-176282)

Qy        195 LeuLeuGlnIleLeuAspValGln------------------AspSerAspAlaGlySer 208
Db     167329 CTCATGGAAACCCTGCCAGTTAAGGGCTAATTCCCCCTTACAGATGGAGACAGCCGGGCT 167270

Qy        209 TyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSerLeuThr 228
Db     167269 CAGAGAGTCCCAGCAAGAGGAGGAAATGAGATTCGGACCAGGACGGTTGTGTTTGCTTT 167210

Qy        229 ValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleValAlaAla 248
Db     167209 ATCCGGGTGCAGAGTGTTCCTGAGGTTCGGAGACTGAAGGGAAGGGCAGTCCCTAGCGTC 167150

Qy        249 ProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAlaSerAla 268
Db     167149 CCCTCCAGC------------------------------------------GCGTCTCCC 167132

Qy        269 AspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThrAspVal 288
Db     167131 TCTCCTTCCCCA--------------------GACGGGAAGCCCATCTCCACAGATGTC 167093

Qy        289 IleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyVal 308
Db     167092 ATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCGCAGCCCTGGCACTCCGGCGTC 167033

Qy        309 TyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeu 328
Db     167032 TATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTCGCCACTGCAGCCGCTGAGCTC 166973

Qy        329 ArgVal------------------------------------------------------ 330
Db     166972 CGTGTGCTGGGTGAGGCGGGCCGAAGGGCGGAGGGAGCGGGTTAGCTGGGGCTGGGGTGA 166913

Qy        330 ------------------------------------------------------------ 330
Db     166912 ATAGAGCAGGGGGAGGTGGACGGGTGGGTGTGCTGGGGGCACTCAGGCGCTCTCCTGGAG 166853

Qy        330 ------------------------------------------------------------ 330
Db     166852 GACGAACTCAGGGAGGAGTCCCTAGTAAGGCGGAGGTGGGACGAGGGATGGGGACTTGGT 166793

Qy        330 ------------------------------------------------------------ 330
Db     166792 AACTGGAGGACGGAGTGCGGGGCGTTGAGTCTGGGATGGAGCGAGAGGGGTGATTAAGGC 166733

Qy        330 ------------------------------------------------------------ 330
Db     166732 CTGGAGCGGGTGATTAGCGGGGAGTTGTGACTGGGGTTATTGTGGGGATAGTGGAGTGAT 166673

Qy        330 ------------------------------------------------------------ 330
Db     166672 CTGAGTGGTACTAAGAAAACTGGGGGGCAGGGTAGATCTGGTAGGTGTGGGACATATGGG 166613

Qy        330 ------------------------------------------------------------ 330
Db     166612 AGGGCAGGGAGGAAGGAGATGGGTAAATTTAGGGGTGGGGTTAGATCCGGGAGGAGCAAT 166553

Qy        330 ------------------------------------------------------------ 330
Db     166552 ATTTAGGGGAGGCTGAGGAGGCCTGGGGACAGTTTTGGGAGAGAAAGAGAATGGTCGTTG 166493

Qy        330 ------------------------------------------------------------ 330
Db     166492 TGGGGCTGGGGTAAATGGGGAGAAGAAAGGGTAGGTCTGCGCTGAGCCGGAGTAGTTTGG 166433

Qy        330 ------------------------------------------------------------ 330
Db     166432 AAAGGGTGAATCAGGGAGAGGTTAAAGTGATTGGAAGGGGAGCGAGGGAGGGCGGTGGGG 166373

Qy        330 ------------------------------------------------------------ 330
Db     166372 GGCAGGCTTGGGGTACTGGGGGGGGGCAGAGACCTTATTTGCGGGATCCCGCGCTCAGC 166313

Qy        331 ----------------LeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerAr 345
Db     166312 GTCGCCCCTCGCGCCCCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCG 166253

Qy        345 gThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLe 365
```

```
Db  166252 GACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCAGCGCT 166193

Qy     365 uHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGl 385
           |   ||||||||:::||| |||||||||||||||||||||||||||||||||||||||||
Db  166192 GCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGG 166133

Qy     385 yGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVa 405
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  166132 CGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGT 166073

Qy     405 lAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGl 425
           ||||||||||||||||||    |||||||||||||||    |||||||||||||||||||
Db  166072 GGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGA 166013

Qy     425 uGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLe 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db  166012 GGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTT 165953

Qy     445 uValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTy 465
           |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db  165952 GGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTA 165893

Qy     465 rGlnLysAlaArgGlyVal----------------------------------------- 471
           ||||||||||||||||||
Db  165892 CCAGAAGGCACGGGGTAC-GTCGGCCTGGGGCGGGATGCACCAGCGCAGGCCCGGGCGGG 165834

Qy     471 ------------------------------------------------------------ 471

Db  165833 GCGCAGCCAGAAGGGGTTCAGGTCGGATGACTGAAGGGTTGCCTGATGGGTACCTCTGCG 165774

Qy     471 ------------------------------------------------------------ 471

Db  165773 GCAAGGATGGACTGGGGTAGGAGCCCATAGGAGAGACCCGCATAAGTATGGTAGGTATGG 165714

Qy     471 ------------------------------------------------------------ 471

Db  165713 ATTAGGCAACTGGCGCTGGAAGGGCCACTGCTTCCCTCAGTCATGTTTGTTGAGAGAGGA 165654

Qy     471 ------------------------------------------------------------ 471

Db  165653 GGGAGAGGACGTTCCTGGAGCCACCTCCCATCCCAGTTAACTAACTGGAGAGGGGGGTCA 165594

Qy     471 ------------------------------------------------------------ 471

Db  165593 TTTCCAGAAAGCACTCCCTAACCAGTGGAGGAATTAATACACTGAGACAGGAAGCTGAGG 165534

Qy     471 ------------------------------------------------------------ 471

Db  165533 GAGTTTATACACTCCAGCTCAGATTATCTGGGATGATTTAGCTATAAATAGACACCAGGA 165474

Qy     471 ------------------------------------------------------------ 471

Db  165473 CAATATGAAGGATTCCTATACAGACTCCAGCTAGCCAGCACGGGGGGGTGGCACTATGAC 165414

Qy     472 -------------------------------AspAsnValGluTyrGlnPheAlaValAs 481
                                          ||||||||||||||||||||||||||||||
Db  165413 AGTCCCCTTGTCCCCCGTGTCACCAGGCATGGACAATGTGGAATACCAGTTTGCAGTGAA 165354

Qy     481 nAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTy 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165353 CAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTA 165294

Qy     501 rValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisTh 521
           |||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db  165293 CGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACAC 165234

Qy     521 rLeuAspAspVal----------------------------------------------- 525
           |||||||||
Db  165233 ACTGGATGATGG-TAGGGCCTCTGAACTCGCAGTGGGCAGCTTGGGGTAAAGAGAATTCC 165175

Qy     525 ------------------------------------------------------------ 525

Db  165174 CTGGGGAGGGGAGCTCACCCTCTCTTTACTAATGTAAAGGCATTTTTACCTGCTCATATC 165115

Qy     525 ------------------------------------------------------------ 525

Db  165114 ATCTTTTGTGCCCCGTAAGAGAGTGGGCGGGGTTCTTCTCCTCAGTTCACAGTTGGGGAA 165055

Qy     525 ------------------------------------------------------------ 525

Db  165054 ACTGAGGCCCAGAAAATGAGAATGATTTGCCCAGGTCAGTAGTTCAGCTGGGACTTAAGC 164995

Qy     525 ------------------------------------------------------------ 525

Db  164994 CCAGGTCTGTGATGATTCCCTGTGTCATGCAGTTCTCCCTGACACCTATGTGCCATCCAG 164935

Qy     525 ------------------------------------------------------------ 525

Db  164934 CTGGGGTAGGGGCCAGTGACATTAACCCATTAGACGTGGAAGATTGTGCTCCAAGACTTT 164875

Qy     525 ------------------------------------------------------------ 525

Db  164874 TTCCAGCTTCAGGGTGGGAAAGATGTCCCCTCATCCTTCAGCTTGGGGCTCCCCTGAGAA 164815

Qy     525 ------------------------------------------------------------ 525

Db  164814 GCTCCCTTTCCTCCCTCAGATCCTGGTCCTGGCCTACTTCTGACTTGCTGTGTGACCCAG 164755

Qy     525 ------------------------------------------------------------ 525

Db  164754 GCAAGTCTCTGACCGCTCCCGAGCCTCACCTTCCACGTCGGTGGAGCAGAGCTCATGCTG 164695

Qy     526 ------------------ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnPr 539
                             |||||||||||||||||||:::||||||||||||||||
Db  164694 ATGTTCTTCTGTCTCAGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCC 164635

Qy     539 oSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLe 559
           ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||::
Db  164634 TTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGT 164575

Qy     559 uLysTyrLysIleGluTyrGlyLeuGlyLysGlu-------------------------- 570
           :|||||||||||||||||||||||||||||||||
Db  164574 GAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGGTGAGTGGGGGTGGGTGGGCTGCAG 164515

Qy     570 ------------------------------------------------------------ 570

Db  164514 GCATGTATGGGAATGATCAACGTGAAGAGCTGTCAGGGCTGCTGGAGACCTGAGCAGGGT 164455

Qy     570 ------------------------------------------------------------ 570

Db  164454 ACGGGAGTCGGAGGAAAGGAAGCAGAGTAGGGGTTGGGGAAGGGTGAACCTGAAAAACAG 164395

Qy     570 ------------------------------------------------------------ 570

Db  164394 GGGTCTGGGCAGGAATGAAAAGGGAAGACAGAGGGGCCTAGCTTAAGGCAGAGGGAGGCC 164335

Qy     570 ------------------------------------------------------------ 570

Db  164334 TGGGCTCTCACCCCAACCCCTCTACAGCTGCAGTGTGTACTTCTACTGAGTCCCTGCCTC 164275

Qy     570 ------------------------------------------------------------ 570

Db  164274 AGTTTCCTTATCTGGGAAATGGGGGGCTGAGGTGGGAGGATCACTTGACCCCAGGAATTC 164215

Qy     570 ------------------------------------------------------------ 570

Db  164214 AAGGCTGCAGGGAACTGTGATCGTGCCATTGTACTCCCATCTGGGTAACAGAGCAAAACT 164155

Qy     570 ------------------------------------------------------------ 570

Db  164154 CTGTCTCAGAAAAAAATAAAAGGATAAAAAGTTAGGAGGAGGAAGGTAGCTGAGGGACCA 164095

Qy     570 ------------------------------------------------------------ 570

Db  164094 GGGGAAGGGTGAGGAAGGAGGGAAAGGTGAGGAAGGAGGGAAGGGTGAGGAAGTGGCCAT 164035
```

```
Qy       570 ------------------------------------------------------------ 570
Db    164034 GGTGGCCCTCAGGGCTCTGGAAGCCAGGCGTTCTTGCTGGCCTTCATGGTCAGCCTCTTG 163975
Qy       570 ------------------------------------------------------------ 570
Db    163974 TGTCTGCATCCCGTGTGTGTGTGTGTTGGCAGGGGATGGTCATTGACAGCTCCTGGGGAA 163915
Qy       570 ------------------------------------------------------------ 570
Db    163914 GTGGGGAAAGCCTGCCATCATTCTCCCTGCAGCTGAGCCCAGGCTGATGGCCTGCATCAG 163855
Qy       570 ------------------------------------------------------------ 570
Db    163854 CCAGAGTTGGAGGCTCAAGCGTCAACCCTGGAGCCATAGAATCTGCCCAGGCACTGAGGG 163795
Qy       570 ------------------------------------------------------------ 570
Db    163794 GTCAGTGCTGAGAACGCCTCAGGCTCCCACGGGCACGGAGGAACCGGGATGTGAGGCGGA 163735
Qy       570 ------------------------------------------------------------ 570
Db    163734 AGAGCCATGTAGGGGGGTTTAGGATCCTGTTCCTCAGGTGGCTACCCAGCAGCCCACCTT 163675
Qy       571 -------------AspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuTh 586
                          |||||:::||||||||||||||    ||||||||||||||||
Db    163674 TTCTGTCCCTGCAGATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACACAGCTTAT 163615
Qy       586 rLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAl 606
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db    163614 GCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGTACAGCAGC 163555
Qy       606 aGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSe 626
             ||||:::|||   ||||||||||||||    |||||||||||    :::|||||||||||
Db    163554 CGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCACAACCAGAG 163495
Qy       626 rHis-------------------------------------------------------- 627
             ||||
Db    163494 CCATGGTACGGGGTTAGAGGAGAGACAGGACAGTGTGTGTGTGACTGTGTGGGTGTGCGT 163435
Qy       627 ------------------------------------------------------------ 627
Db    163434 GTGTCCTTGGGACCCAGCGCCCCACCCCCTGGCCCCGCCCTCCATCAGAGACCTTCCATG 163375
Qy       627 ------------------------------------------------------------ 627
Db    163374 GCTCCCTCCGCTTGATGGGTGAGTGAGGCTTAGGTGCAGAGGTGAAGGGCTGGGATCTGG 163315
Qy       627 ------------------------------------------------------------ 627
Db    163314 TGTCCCTTCGATTTGCTGTGTGACCTTGGACAAACCATTCTGAGGCTTCATTTACTGCCC 163255
Qy       627 ------------------------------------------------------------ 627
Db    163254 AGGAATGGCATGAGGTTGAATAATGTATTGTAACTAAAGTGCTTAGTACATAAAAACTCA 163195
Qy       627 ------------------------------------------------------------ 627
Db    163194 TTGAATTGTAGCTGCTTCGAGGATGATGGTGATGATGATTGTTTTTGTTGAGAGGCCCAA 163135
Qy       627 ------------------------------------------------------------ 627
Db    163134 TAGAAGAGTGATTAAATGCTGACCAAACCGGTCTACATGCTCTCTATGCGACACTAGGTA 163075
Qy       627 ------------------------------------------------------------ 627
Db    163074 AATTACTTAACTCTCTGAGCTTCAGTTTTCTTATCTGTAAAGCGAAGGTTGTAGTATTTT 163015
Qy       627 ------------------------------------------------------------ 627
Db    163014 CACTATTTTAGAGGGTTGCAAGGATAAAGTTATAATAACTACACAAAGTGCATAGGACAG 162955
Qy       627 ------------------------------------------------------------ 627

Db    162954 TGTTTGACACATTGTAAGCTCCTCACTCTCAAGTGGGTCTCCCTACTGTAATGGTGGCAT 162895
Qy       627 ------------------------------------------------------------ 627
Db    162894 CCCTCCTTCCTGGCCTCAGCAACCACCACCCCCACTGCCCCCAGTGTTTGATCCTTCAGA 162835
Qy       627 ------------------------------------------------------------ 627
Db    162834 AGACTTCACAGAGGCCGATTCTGTGCCAGGTGCTATCAGCCTACTCCAGAGGTGCTCCTT 162775
Qy       627 ------------------------------------------------------------ 627
Db    162774 GCTGACCCCTTTGCAAATGCCCAGCGGACACCCCTGCATTCCCTGCACTCTGTGCTTGTA 162715
Qy       627 ------------------------------------------------------------ 627
Db    162714 GAACAGTGAGTTCACCTCTAGACCAAGCCCAGGACTGGCACTTAGAGTGGTGGCTGAGTG 162655
Qy       627 ------------------------------------------------------------ 627
Db    162654 GTGATGTGTTGGGAGAATGCTGGCAAAAGGCTCTCACTGAATGCCATCTTCCTGTCTCTT 162595
Qy       627 ------------------------------------------------------------ 627
Db    162594 CCTCATGCCTCTTCCCTGTCTCCTATCCCTGGCCTCCTTCTCCTCATTCTGCATTCTGTT 162535
Qy       628 ---------------------------------ValProPheAlaProAlaGluLeuLy 636
                                              |||||||||||||||||||||||||||
Db    162534 CCTTAATTTTCCCATCCTGCTCCTCTGCCCCCCAGTCCCTTTTGCCCCTGCAGAGTTGAA 162475
Qy       636 sValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGl 656
             ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    162474 GGTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCA 162415
Qy       656 nIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluAlaAspGlyAs 676
             |||||||||||||||||||||||||||    |||||||||    ||||||||||:::||||
Db    162414 GATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGA 162355
Qy       676 pArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLy 696
             ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    162354 TCGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAA 162295
Qy       696 sLysValLysGlnTyrGluLeuThrGlnLeu----------------------------- 706
             |||||||||||||||||||||||||||||||||
Db    162294 GAAAGTGAAGCAGTATGAGCTGACCCAGCT-AGGTGAGGAGGGCTGGGGGGAGGGGGAGG 162236
Qy       706 ------------------------------------------------------------ 706
Db    162235 GAGGGATGTAGCTTCAGTGGGGACTGGGCAAGTTGTGTGAAGAAGTTAGAGATGGGATGC 162176
Qy       706 ------------------------------------------------------------ 706
Db    162175 TCTCTGTGGACGGGAGAGTGACCCCAGATGTCCCCCTAGCCATTAAGCCCTGTAACCTTC 162116
Qy       707 ----ValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGl 725
                 |||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db    162115 CCCAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAGGATGG 162056
Qy       725 yTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAsp----------- 741
             ||||||||||||||||||||||||||||||||||||||||||||||    |||
Db    162055 CTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGG-TGAGGGCGGTG 161997
Qy       741 ------------------------------------------------------------ 741
Db    161996 GGGGAAGAAGGCGGGGAGGGCTCAGGTGAGGGCTCTGCTAGCCGCTTCCTCTTCTTTTCC 161937
Qy       741 ------------------------------------------------------------ 741
Db    161936 TTTGGGCCAGCCTGGTGAGCAGGTACTTTAGTAGCCTCAGCTCATGCCCAAACCCATTCT 161877
Qy       741 ------------------------------------------------------------ 741
```

```
Db  161876 CCTAGGATCAGGGGTCCTCTTTTAGCTCCAGAGAGCCTCTGTCATGGAGAGAAGCTTGTT 161817
Qy     741 ------------------------------------------------------------ 741
Db  161816 TGTTCACTGTCTTAGGCCTGCCACCTTATTTGCATTACCCATGAATAGTGCTGATTTCCA 161757
Qy     741 ------------------------------------------------------------ 741
Db  161756 AAGTTTGAACAGAGGGAAACCCTATGGTTGATCAGTAACTCCCGACCCCAAGTTTGACTG 161697
Qy     741 ------------------------------------------------------------ 741
Db  161696 TAAAACAAGTTTTTTTAAAAACACATTTGTCTACATCACACTATCAGACTATCTCTCAAG 161637
Qy     741 ------------------------------------------------------------ 741
Db  161636 TCTAAGAAGAGTGAGTATCCCTGTGTGACCTAACGGGAAACTGAGGCTCAGAGAGGAAAA 161577
Qy     741 ------------------------------------------------------------ 741
Db  161576 GGAACTTGTCCAAGGTCACACAGAGCCTAGGCTACAACTGGGACTCCTGGTTTCCAGGCA 161517
Qy     741 ------------------------------------------------------------ 741
Db  161516 AAACTCTTTCCACTGCACTACACTGCCTTGTGGCCCTGCATTGGACACCTCAGTTGCCCA 161457
Qy     741 ------------------------------------------------------------ 741
Db  161456 TCATGATACTCCCTCACCCCTAACTTCCCAGCCACATGGTTTAGCAGAGGGAGTAATCAA 161397
Qy     741 ------------------------------------------------------------ 741
Db  161396 CCAGAAGCCATGAAGCTTGGGTTCTTTTTTTTTCTTTTGAGACAGAGTCTCACTCTGTTG 161337
Qy     741 ------------------------------------------------------------ 741
Db  161336 CCCAGGCTGGAGTGCAATGGTATGATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCA 161277
Qy     741 ------------------------------------------------------------ 741
Db  161276 AGCAATTCTTGTGCCTCGGCCTCTTGAGCAGCTGGGATTACAGGTGTGCACCATACCTGG 161217
Qy     741 ------------------------------------------------------------ 741
Db  161216 CTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCCTTGAA 161157
Qy     741 ------------------------------------------------------------ 741
Db  161156 CTCCCGACCTCAGGTGATCCACCCATCTTGGCCTCCCAACGTGCTGGGATTACAGGTGTG 161097
Qy     741 ------------------------------------------------------------ 741
Db  161096 AGCCACCACACCCAGCTGAGTTCTAATTCCAGTTTGTTCATTTCTTCATTCAGCAAATAT 161037
Qy     741 ------------------------------------------------------------ 741
Db  161036 TTCGCTGAGTGTCTACTAAGTGCCAGCTGCTTTGGGAATTTAGTGGTGGACAACACAAAT 160977
Qy     741 ------------------------------------------------------------ 741
Db  160976 AAGGTTCTTGCTCTCTTGGGGTTTATTTTATGTTTCTAAAATAAACACTTATAAATATAG 160917
Qy     741 ------------------------------------------------------------ 741
Db  160916 TGTTTATTATATTTCAATTTCCTATCTATTTTTAAAATATTAAATAATTTAATCCTCATA 160857
Qy     741 ------------------------------------------------------------ 741
Db  160856 ACAACCCCATGAGGTAGGTACTATCATTAGAATAGTGAATGGAGACACAGAGAAGGTAAG 160797
Qy     741 ------------------------------------------------------------ 741
Db  160796 TAATTTCTCCAAGGTCACACAGCTAGTAAGTGTCAGAGCTGGAATTTGAACCAAGGCAGC 160737

Qy     741 ------------------------------------------------------------ 741
Db  160736 CTGGTTCCAGAGTCTGGTTCCATGGTTATTCATTGCTGTGATTTGCTGCCTCTCTCTTTA 160677
Qy     741 ------------------------------------------------------------ 741
Db  160676 TTACGTTCTTACAGACTAGTGTTCCACAGTCAAAAATCAAATTCTTGGAAGCCTCAATTG 160617
Qy     741 ------------------------------------------------------------ 741
Db  160616 CCTTATCTGTGAAGTAGGAATGATGCCACTCAACTCACAGGATCAAAGGAGTGGGTCAAA 160557
Qy     741 ------------------------------------------------------------ 741
Db  160556 GAAACAGCCTTGGGAACTGTAAACCACTATACATATTTAAGGATGAGTATCCTCCCTCCC 160497
Qy     742 -------------------------------LeuProIleGlnArgGlyProProLeuPr 751
                                          :::||||||||||||||||||||||||||
Db  160496 TTCCTCACACTCTAGTTTTCACCTTTCAGACATGCCTATCCAGAGGGGACCACCCCTGCC 160437
Qy     751 oProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLy 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  160436 TCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAA 160377
Qy     771 sProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLe 791
           |||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||
Db  160376 GCCAGATTTCACCACAGTCAAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCT 160317
Qy     791 uArgAsnAlaSerLeuValThrTyrTyrThrSer-------------------------- 802
           |||||||||||||||||||||||||||||||||
Db  160316 CAGGAATGCCTCCCTGGTCACCTATTACACCAG-GTGGGGACTAATGCCTTTACTAATGT 160258
Qy     802 ------------------------------------------------------------ 802
Db  160257 AAAGAAGGGAGGAAGAATAGGTGGGGGGTGGTGGTGGGAAGGCAGCAGGTGGTGATCAGG 160198
Qy     802 ------------------------------------------------------------ 802
Db  160197 GAAGGGCCCCTGGGATTCAGGAACATGTCATGGCTCACCTGTTACGCTGAGGCCACACAG 160138
Qy     802 ------------------------------------------------------------ 802
Db  160137 AGGGGCTGCACAGGGTCCACGTCCTTGCCTGCTGACCAGATGGTAGCAAGGAGGCCATCC 160078
Qy     802 ------------------------------------------------------------ 802
Db  160077 AATCCCGTCCCCTCACAACAGAGCCTTGTTCTCAAACTCCGGAGAATGTGACACAGGACA 160018
Qy     802 ------------------------------------------------------------ 802
Db  160017 GACCTGTTTCTCCAACCTCAGAGGGCTCCCAGTCTGATGGTGAAGAGAACCCCACCCCTG 159958
Qy     802 ------------------------------------------------------------ 802
Db  159957 TCCTCAGAGCTGCTCCCTCTACCCTGTTAGATGGGGAGACAGAGCCTGAGGAGCCTGAGG 159898
Qy     802 ------------------------------------------------------------ 802
Db  159897 TTTGGCGGAGGAAAGTCGCATAAAAAGCAACCGATATATACAACTGCCATGAAGGCAAAA 159838
Qy     802 ------------------------------------------------------------ 802
Db  159837 TTAAGCCAAAGAAAAGATGCATCGGAGATGATTCCATAGGTGGAGAATAACTTGGATTAA 159778
Qy     802 ------------------------------------------------------------ 802
Db  159777 TCCAGGAAGGTTTCCTGGAGGAAGGGGTCTGGTTTCGAGATTGGACTTGGAGCAGCATAC 159718
Qy     802 ------------------------------------------------------------ 802
Db  159717 ATCCAGACTTCTCTGTAGGTCCCTGGGTGGTGGTAGTCGCCTCCAGGCCAGGGAGCAGCT 159658
```

```
Qy        802 ------------------------------------------------------------ 802
Db     159657 TCCCTTCCCAGGGTAGAGCCTGGGGCTGGGGCTGACTGCAGGGGAGTAGATTTGGAGGAG 159598
Qy        803 ----------------------------------------------------------Se 803
                                                                        ||
Db     159597 TCTGCCCTGCTGGGAGAGTGTGCCTGAGAGCACTAGTCACGTTTCTTGCCTTAGCAGTTC 159538
Qy        803 rGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGl 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159537 TGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCA 159478
Qy        823 nSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThrLeuPr 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159477 GTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCC 159418
Qy        843 oAspArg----------------------------------------------------- 845
              |||||||
Db     159417 TGACCGTGAGTGGCTCCCAGCCCCTTGCCACCAACCCATCTTTTGCCCAACTCCTTCTTG 159358
Qy        845 ------------------------------------------------------------ 845
Db     159357 CCCCAGCAGCAAGAAACATACTTCAGAACCCCAAGGTTTGGGCAGAGCTGGGGGCAGAAG 159298
Qy        845 ------------------------------------------------------------ 845
Db     159297 AGTTTCTGGGAGGGGAGGAATGTTAAGTCCCTCCAACCACACATACTGTTTTCTAGAATT 159238
Qy        845 ------------------------------------------------------------ 845
Db     159237 CCCCAACTAGAAACCAGACCTCATCTTTGCTTTGTGGCTTGTTCCATGAAGGAGGAAGAA 159178
Qy        845 ------------------------------------------------------------ 845
Db     159177 GGGAGGTCATATCATCATGCTCTGTCCCCCAGCTCTGACATCTCTCCCCATCTCCTCTCC 159118
Qy        846 -------------ProSerThrProProSerAspLeuArgLeuSerProLeuThrProSe 861
                           |||  ||||||||||||||||||||||||||||||||||||||||||
Db     159117 CGCACCCCAGGGCCCTCC-ACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTC 159059
Qy        861 rThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLe 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159058 CACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCT 158999
Qy        881 uIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGl 901
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db     158998 GATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCA 158939
Qy        901 u----------------------------------------------------------- 901
Db     158938 -GGGTGAGGGCTCCTGCCCCAGGGCATCTGCTCAAGAGCTTCCCAGCCTCTCCCGCTTCC 158880
Qy        901 ------------------------------------------------------------ 901
Db     158879 AGCAGTGAGAACAGTGCTGGGGTTTGTTCCAGGGGCTACTGGGAGAAGCAAGCAGAAGCT 158820
Qy        901 ------------------------------------------------------------ 901
Db     158819 GCTGCCTAAGTGGAGAGCAGACTCTCCCAGTCCCCTCCCTGCCCACCTCTGCATCCACCC 158760
Qy        902 -------------------------------------------GlyAsnIlePheSerAl 907
                                                         |||||||||||||||||
Db     158759 ACCCCATCCACCATCCCTGCTCTCCCTGCCTCCCTCACCCCCAGGAAACATCTTCAGTGC 158700
Qy        907 aGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGl 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158699 TGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGA 158640
Qy        927 uValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSe 947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db     158639 GGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTC 158580
Qy        947 r----------------------------------------------------------- 947

              |
Db     158579 AGGTATGAGCACAGGGCAGCCCAGGTTTCCCTGGATTGACTTCCAAGCCCCTGCCTGTGG 158520
Qy        947 ------------------------------------------------------------ 947
Db     158519 ACACAACTACAATTTCATTGTTTTGCACAGCTGTCCTCCAGGTTTGGTGGGGGCAGAGCA 158460
Qy        947 ------------------------------------------------------------ 947
Db     158459 AATGTGAAGGGGAAATAGGTGTGGTTCAGTGATAACTTGATTCAGAGGCCAGGTTCGTGG 158400
Qy        947 ------------------------------------------------------------ 947
Db     158399 GGGACTGGCTTTTTTTTTTTTTTTTCTGAGATGGAGTTTCACTCTTGTTGCCCAGGCTAGA 158340
Qy        947 ------------------------------------------------------------ 947
Db     158339 GTGCAATGGCGCAATCTCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCGATTCTGC 158280
Qy        947 ------------------------------------------------------------ 947
Db     158279 TGCCTCAGCCTCCCAAATAGCTGGGATTACAGGCACACACCACCATGCCCGGCTAATTTT 158220
Qy        947 ------------------------------------------------------------ 947
Db     158219 TGTATTTTTAGTAGAGACAGGATTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 158160
Qy        947 ------------------------------------------------------------ 947
Db     158159 CTCGTGATCTGCCTACCTTGGCCCCCTAAAGTGCTGGGATTACAGGCGTGAGCCACCATG 158100
Qy        947 ------------------------------------------------------------ 947
Db     158099 CCCAGCCAATTATTGTATTTTTTAGTAAAGACGGGTTTTCACCACATTGGCCAGGCTGGT 158040
Qy        947 ------------------------------------------------------------ 947
Db     158039 CTCAAACTCCTTACCTCAGGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACA 157980
Qy        947 ------------------------------------------------------------ 947
Db     157979 GGTGTGAGCCACTGCATCCAGCTGGGACTGGCTTTACAGACCAGCTTCTTTGGGAAATTG 157920
Qy        947 ------------------------------------------------------------ 947
Db     157919 CTTTGATTGGAGGATTAACTTGGTGAGGTGATGAGCTTTATGGTGTGGAGGTGGACAGCA 157860
Qy        947 ------------------------------------------------------------ 947
Db     157859 GGACTGGCTACGTAATTTCCAGGGCCCAGTGCAGAATGAAAAGGTGCTATCCCTTATTTT 157800
Qy        947 ------------------------------------------------------------ 947
Db     157799 AAAAATTAGAGGTCGGACGGCCAGGCGCGGTGGCTTATGCCTGTAATCCCAGCACTTTGG 157740
Qy        947 ------------------------------------------------------------ 947
Db     157739 GAGGCCAAGGTGCGGGCGGATCATGATGTCAGGAGATCGAGACCATCCTGGCTAACACGG 157680
Qy        947 ------------------------------------------------------------ 947
Db     157679 TGAAACTCTGTCTCTACTAAAAATACAAAAAAAAATTAGCTGGGCGTGGTGGTGGATGCC 157620
Qy        947 ------------------------------------------------------------ 947
Db     157619 TGTAGTCCCAACTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGAGGAG 157560
Qy        947 ------------------------------------------------------------ 947
Db     157559 CTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCCG 157500
Qy        947 ------------------------------------------------------------ 947
```

```
Db  157499 TCTCAAAAACAAAAACAAAAACAAAAAAAGCTAGAGGTCGGGCATGGTGGCTTGTGCTTG 157440

Qy     947 ------------------------------------------------------------ 947

Db  157439 TAATCCAAAGTGCTTATGCACTTTGGAAGGCTGAGGTGAGAGGATCACTTGAGGCCAGGA 157380

Qy     947 ------------------------------------------------------------ 947

Db  157379 GTTGGAGACCAGCCTGGACAACATGGTGAGACCCTGTCTCCACAAAAAATTTAAAAATTA 157320

Qy     947 ------------------------------------------------------------ 947

Db  157319 TCTGGGTGTGGTAGCTCATGCCTGTGGTCACAGCTGCTTGGGAGGCTGAGGTGGGAGGAT 157260

Qy     947 ------------------------------------------------------------ 947

Db  157259 CACTTGAGCCCAGGAGGTTGATGCTACAGCGAGCCGTGTTCATGCCACTGCACTCCGGCC 157200

Qy     947 ------------------------------------------------------------ 947

Db  157199 TGGGTGACAGAGCAAGACCCTGTCTCAAAAGAAAAAAAATTAGAATTTTAAGGTGGTGAC 157140

Qy     947 ------------------------------------------------------------ 947

Db  157139 AGCAGAGCATGAACTCAAGCGTGGGGTCCATCCAAGCATGGGGTCTGCGCACCTGCACAT 157080

Qy     947 ------------------------------------------------------------ 947

Db  157079 GTTGCACACCCATGCAACAGCCCTGGGAGCAGGCAACACGTGCTTCACCAAAGGCCAGTT 157020

Qy     947 ------------------------------------------------------------ 947

Db  157019 TGAGGCAGAGAGACCATAACTTCTGGGTGTGTTTTAGGGAGGACAGTCAGAGGGAGAGGT 156960

Qy     947 ------------------------------------------------------------ 947

Db  156959 AGTAGTAGTTTCTGTGAGGCTCAGGGCCCTCCTGTCCTTGTATCAGAAGATTGTGAGCAT 156900

Qy     947 ------------------------------------------------------------ 947

Db  156899 CTGAAGGGTGAGGGTCCAGCTGGTGTCTCCTGAGAAAGCCATGGATAGTCCTGATCTATG 156840

Qy     948 -----------------------------------------------------AspSerL 950
                                                                |||||||
Db  156839 CCCTGTGCCCTTCCTCTGTCCTGACCTGCTCTGTGCCTCTCTCTCCCCCACCAGACTCGC 156780

Qy     950 euAspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuL 970
           ||||:::|||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db  156779 TGGACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCC 156720

Qy     970 euAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArg------------------- 983
           ||||||||||||||||||||||||||||:::|||   |||
Db  156719 TGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAG-GTGAGCAGGGGAGCCAGTC 156661

Qy     983 ------------------------------------------------------------ 983

Db  156660 TGACTTCAGGGAAGGGAGACTTCCAGGAAGGGCTGCAGAATGGGCCATCAGGGAAAGTGG 156601

Qy     983 ------------------------------------------------------------ 983

Db  156600 CCATGACTTGCCCTTGGCACTCCCAGGAGCCCCGTTAAAGTCAGGACTCTGGGATGAAGG 156541

Qy     983 ------------------------------------------------------------ 983

Db  156540 AATCAGGGATGATGTCCTCCTGAGAATGGGGAGTGGACTTAGGGGAGCCCCAGCCTACCC 156481

Qy     983 ------------------------------------------------------------ 983

Db  156480 TCCTCAGCAAGGCCAACAGAACTGCCTCACTCCCACTTCCTCTGTAGCTGCCAGCATCTT 156421

Qy     983 ------------------------------------------------------------ 983

Db  156420 GGAAAGAGGGGCCCCAAATCTCCAGATGCCTGCAGCCAAGGGAAGAAGAGCCCTGTTTAC 156361

Qy     983 ------------------------------------------------------------ 983

Db  156360 AGCATCAGGCTGTTTTGAACAATAGCAGTTTTTCTTACTCATTGCACAGACCAGGAAAGC 156301

Qy     983 ------------------------------------------------------------ 983

Db  156300 AAATCCCAGAGATGGGCAGGAACCTCCCCAGAGACATAGGCTGGACCAGATCCCAGCCCC 156241

Qy     984 ----------------------------------------------------GluAlaL 986
                                                               |||:::|
Db  156240 GCCTGACCCCTTACTCTGACCGCTATGCCTTGTTTTCCGTCTTCCCTTCCAGGGAATCCC 156181

Qy     986 euProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgL 1006
           ||||||||||||||||:::   ||||||||||||||||||||||||:::||||||||||
Db  156180 TCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGAGCTCGGC 156121

Qy    1006 euGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisProArgProG 1026
           ||||||||||||||   |||||||||||||||||||||||||||||||||||||   ||||
Db  156120 TTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCCCATCCCC 156061

Qy    1026 lnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSerLeuMetG 1046
           |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db  156060 AGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGCCTTATGG 156001

Qy    1046 lyGlySerValSerAspCysArgGlyHisSerLysArgLys------------------- 1059
           |||||   ||||||:::    |||    ||||||||||||||
Db  156000 GTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGGTAAGTCAGAGCCCCACAT 155941

Qy    1059 ------------------------------------------------------------ 1059

Db  155940 TGCGGCTCTTTTCTAAATAGGACACCAGGGGGCGACAGAGCACAACGCGTGGCTATGCGA 155881

Qy    1059 ------------------------------------------------------------ 1059

Db  155880 GCATGGCTCTACCTCCTCTTCGGAGCCGGCTGCGGACGCTAGGGCCTCTCTCCCTCCTCC 155821

Qy    1059 ------------------------------------------------------------ 1059

Db  155820 ACGGAATGGTTGGCTGTCAGGGAAGCACATGGGCCTCCATATCCCGGCGGAATAGGTCAG 155761

Qy    1059 ------------------------------------------------------------ 1059

Db  155760 CTTGGTTCTCACGAAGGAAAGGGCCTCCCAGCAGGTGACCCAGGGTAGCCAGAACACCCA 155701

Qy    1059 ------------------------------------------------------------ 1059

Db  155700 GCTCCCTCCTCATCCCTGGAGCTGGGGAGACCTCTGAGACCCAGACATCCCCGTGTGGGG 155641

Qy    1059 ------------------------------------------------------------ 1059

Db  155640 AGAAGAGGCCAGTCGGCTTGGCTGTGGCCAAGTGGGACTGGGACCAGGATAAGGTGAGGC 155581

Qy    1059 ------------------------------------------------------------ 1059

Db  155580 CTTCGCTGGGCTGGTCCCTTTCATGCTGGCAGGGGCGGCACATCCCACCCAACTCCGACC 155521

Qy    1059 ------------------------------------------------------------ 1059

Db  155520 CGTCCTCCCTCACCCTCTCCCCGCCTGGCTCGGACACCGCGGGCTTTGCAGAGCGTCTGG 155461

Qy    1059 ------------------------------------------------------------ 1059

Db  155460 TGTTCCGCTCCTCTGCAGGAGGACTCATCCTCAGATTACCCCGCTGACTCTGAGCCCTTT 155401

Qy    1059 ------------------------------------------------------------ 1059

Db  155400 CGATGCTCCTCCCCAGAGCCTTGGAGAGGGAGAGAGCAGCGGGTGACAGTCTGGACCCTG 155341

Qy    1059 ------------------------------------------------------------ 1059

Db  155340 CTCCCCCGCGGCCCATCCCAATTCCCTCCCCCGGGCTTCCCTCTCCCAGCCAATCCTGGT 155281
```

```
Qy       1060 -----------------------------------------IleSerTrpAlaGlnAlaGlyGly 1067
                                                         ||||||||||||||      |||
Db     155280 CGTGCCCCTTACTGTCCCCTTCTTTCCCCCGTGCAGATCTCCTGGGCTCAACCAAGCGGG 155221

Qy       1068 ProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerGlyProArgPro 1087
                 :::||||||||||||||||||||||||||||||||||||||      :::||||||||||||
Db     155220 CTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAG---GCAGGCCCCCGGCCG 155164

Qy       1088 AlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAla 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155163 GCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGCT 155104

Qy       1108 LeuValTyrAspGlyIleLys--------------------------------------- 1114
              ||||||||||||    ||||||
Db     155103 CTGGTGTACGACGCCATAAAGGTCAGCGTCCTGGAAGGAAGGGGCAGGAAAGGAGGAAAG 155044

Qy       1114 ------------------------------------------------------------ 1114

Db     155043 GGTCTCTGTGCTTGTGAAGAGGAAGTAATGAGTTCTCTGGAGACCCCTGGGTGGTAACTT 154984

Qy       1114 ------------------------------------------------------------ 1114

Db     154983 TCCTGGGGATAGAAGTCTAGCCAGGAGACCCCTGGAGCCTCCAAAGGTTGAGTAGGTGGA 154924

Qy       1114 ------------------------------------------------------------ 1114

Db     154923 GGGCAGGCCAGTTTGCAGGAACCTGTTTCAACCTGGGTTGAAAACAGCCCCTCCTCTACC 154864

Qy       1114 ------------------------------------------------------------ 1114

Db     154863 CAGACCTAAGTGGGCCACAGCAGTTCCGTGGATCTGAGGGTGTGGCTGGCCTGAGTACGC 154804

Qy       1114 ------------------------------------------------------------ 1114

Db     154803 TCAGCCCTTCCTCCTGTCCCCAGAGGCCCTGACTGAGTCTGGTACTCCTGAAACTCCTGG 154744

Qy       1114 ------------------------------------------------------------ 1114

Db     154743 ATGAGGGCAAATGGGAAGTATTCCATCATTCCCTGCCTTTGTATAGGTTCGAACAGTTCC 154684

Qy       1114 ------------------------------------------------------------ 1114

Db     154683 TGTGTATCCTTCACCGACTCCCATGTGATACTTACTCTGATGGAGGGTCAGACTTTCTGC 154624

Qy       1114 ------------------------------------------------------------ 1114

Db     154623 CCCACGTACGCTGGTCATTCTGATGATTACTCTCAGTAATGGTCACTTTGTCTTTGTGTT 154564

Qy       1115 ---------------SerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnVal 1129
                               |||||||||||||||      ||||||||||||||||||||
Db     154563 CTCTGCTCTGATCAGGGCAATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTG 154504

Qy       1130 GluAlaGluValIleValHisSerAspPheGlyAlaSerLysGlyCysProAspLeuHis 1149
              |||||||||||||||||||||||||||||||    ||||||   |||   |||||||||||
Db     154503 GAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCAT 154444

Qy       1150 LeuGlnAspLeuGluProGluGluProLeuThrAlaGluThrLeuProSerThrSerGly 1169
              |||||||||||||||||||||:::||||||       |||             ||||||
Db     154443 CTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGT 154384

Qy       1170 AlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnPro 1189
                 |||      |||||||||     ||||||   ||||||||||||||||||:::
Db     154383 GTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTG 154324

Qy       1190 ThrThrSerGlyProGluArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSer 1209
                       ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db     154323 GCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCC 154264

Qy       1210 CysSerAspLeuGlnProSerThrAlaIleGluGluAlaProGlyLysSerCysGlnPro 1229
                 ||||||||||||          :::||||||   ||||||   |||||||||
Db     154263 TACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTC 154204

Qy       1230 LysAlaLeuCysProLeuThrValSerProSerLeuProArgAlaProValSerSerAla 1249
              |||:::   |||||||||      ||||||     |||||||||:::||||||||||||:::
Db     154203 AAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 154144
```

```
RESULT 14
AC105129
LOCUS       AC105129               198295 bp    DNA     linear   PRI 26-APR-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-349G13, complete sequence.
ACCESSION   AC105129
VERSION     AC105129.4  GI:20330981
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
```

```
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 26, 2002 this sequence version replaced gi:20148056.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22971
                Center clone name: 349_G_13
            --------------.
FEATURES             Location/Qualifiers
     source          1..198295
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1..79)
                     /rpt_family="AluS"
     repeat_region   complement(183..494)
                     /rpt_family="AluSq"
     repeat_region   complement(1216..1416)
                     /rpt_family="L1MB8"
     repeat_region   complement(1417..1709)
                     /rpt_family="AluSg"
     repeat_region   complement(1710..2115)
                     /rpt_family="L1MB8"
     repeat_region   2116..2169
                     /rpt_family="(TTTTA)n"
     repeat_region   2186..2234
                     /rpt_family="AluSq/x"
     repeat_region   complement(2236..2512)
                     /rpt_family="AluJb"
     repeat_region   complement(2514..2631)
                     /rpt_family="L1MB8"
     repeat_region   complement(2632..2941)
                     /rpt_family="AluY"
     repeat_region   complement(2942..3104)
                     /rpt_family="L1MB8"
     repeat_region   3181..3464
                     /rpt_family="AluSx"
     repeat_region   3465..3711
                     /rpt_family="AluJb"
     repeat_region   3712..3788
                     /rpt_family="(TA)n"
     repeat_region   complement(4661..4940)
                     /rpt_family="AluJb"
     repeat_region   complement(4946..5106)
                     /rpt_family="FRAM"
     repeat_region   complement(5878..6193)
                     /rpt_family="AluYb8"
     repeat_region   7197..7490
                     /rpt_family="AluSx"
     repeat_region   7925..8226
                     /rpt_family="AluY"
     repeat_region   complement(8227..8312)
                     /rpt_family="L1MB3"
     repeat_region   8313..8347
                     /rpt_family="AT_rich"
     repeat_region   complement(8350..8617)
                     /rpt_family="AluSx"
     repeat_region   complement(8621..9274)
                     /rpt_family="L1MB3"
     repeat_region   9403..9445
                     /rpt_family="AT_rich"
     repeat_region   9706..10007
                     /rpt_family="AluJb"
     repeat_region   complement(10940..11000)
                     /rpt_family="L1MB5"
     repeat_region   11001..11284
                     /rpt_family="AluJo"
     repeat_region   complement(11285..11366)
                     /rpt_family="L1MB5"
     repeat_region   complement(11641..11934)
                     /rpt_family="AluSg"
     repeat_region   complement(12071..12294)
                     /rpt_family="AluJb"
     repeat_region   complement(12410..12697)
                     /rpt_family="AluSx"
     repeat_region   complement(12708..12998)
                     /rpt_family="AluJb"
     repeat_region   13158..13478
                     /rpt_family="AluJo"
     repeat_region   13970..14093
                     /rpt_family="FLAM_C"
     repeat_region   14108..14412
                     /rpt_family="AluJo"
     repeat_region   complement(14813..15119)
                     /rpt_family="AluSx"
     repeat_region   complement(15130..15432)
                     /rpt_family="AluSq"
     repeat_region   complement(15532..15841)
                     /rpt_family="AluSx"
     repeat_region   complement(15906..16077)
                     /rpt_family="AluSg/x"
     repeat_region   complement(16079..16383)
                     /rpt_family="AluSq"
     repeat_region   16409..16481
                     /rpt_family="MER5A"
     repeat_region   16482..16779
                     /rpt_family="AluSq"
     repeat_region   16780..16874
                     /rpt_family="MER5A"
     repeat_region   complement(17247..17543)

Alignment Scores:
Pred. No.:              1.06e-93      Length:         198295
Score:                  2557.00       Matches:        877
Percent Similarity:     20.85%        Conservative:   45
Best Local Similarity:  19.83%        Mismatches:     111
Query Match:            38.72%        Indels:         3397
DB:                     9             Gaps:           18

US-09-754-997A-2 (1-1252) x AC105129 (1-198295)

Qy        195 LeuLeuGlnIleLeuAspValGln------------------AspSerAspAlaGlySer 208
              |||::::::  |||   |||:::                   |||   |||:::   :::
Db      88556 CTCATGGAAACCCTGCCAGTTAAGGGCTAATTCCCCCTTACAGATGGAGACAGCCGGGCT 88615
```

```
Qy      209 TyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSerLeuThr 228
Db    88616 CAGAGAGTCCCAGCAAGAGGAGGAAATGAGATTCGGACCAGGGACGGTTGTGTTTGCTTT 88675
Qy      229 ValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleValAlaAla 248
Db    88676 ATCCGGGTGCAGAGTGTTCCTGAGGTTCGGAGACTGAAGGGAAGGGCAGTCCCTAGCGTC 88735
Qy      249 ProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAlaSerAla 268
Db    88736 CCCTCCAGC------------------------------------------GCGTCTCCC 88753
Qy      269 AspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThrAspVal 288
Db    88754 TCTCCTTCCCCA---------------------GACGGGAAGCCCATCTCCACAGATGTC 88792
Qy      289 IleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyVal 308
Db    88793 ATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCGCAGCCCTGGCACTCCGGCGTC 88852
Qy      309 TyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeu 328
Db    88853 TATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTCGCCACTGCAGCCGCTGAGCTC 88912
Qy      329 ArgVal------------------------------------------------------ 330
Db    88913 CGTGTGCTGGGTGAGGCGGGCCGAAGGGCGGAGGGAGCGGGTTAGCTGGGGCTGGGGTGA 88972
Qy      330 ------------------------------------------------------------ 330
Db    88973 ATAGAGCAGGGGGAGGTGGACGGGTGGGTGTGCTGGGGGCACTCAGGCGCTCTCCTGGAG 89032
Qy      330 ------------------------------------------------------------ 330
Db    89033 GACGAACTCAGGGAGGAGTCCCTAGTAAGGCGGAGGTGGGACGAGGGATGGGGACTTGGT 89092
Qy      330 ------------------------------------------------------------ 330
Db    89093 AACTGGAGGACGGAGTGCGGGGCGTTGAGTCTGGGATGGAGCGAGAGGGGTGATTAAGGC 89152
Qy      330 ------------------------------------------------------------ 330
Db    89153 CTGGAGCGGGTGATTAGCGGGGAGTTGTGACTGGGGTTATTGTGGGGATAGTGGAGTGAT 89212
Qy      330 ------------------------------------------------------------ 330
Db    89213 CTGAGTGGTACTAAGAAAACTGGGGGGCAGGGTAGATCTGGTAGGTGTGGGACATATGGG 89272
Qy      330 ------------------------------------------------------------ 330
Db    89273 AGGGCAGGGAGGAAGGAGATGGGTAAATTTAGGGGTGGGGTTAGATCCGGGAGGAGCAAT 89332
Qy      330 ------------------------------------------------------------ 330
Db    89333 ATTTAGGGGAGGCTGAGGAGGCCTGGGGACAGTTTTGGGAGAGAAAGAGAATGGTCGTTG 89392
Qy      330 ------------------------------------------------------------ 330
Db    89393 TGGGGCTGGGGTAAATGGGGAGAAGAAAGGGTAGGTCTGCGCTGAGCCGGAGTAGTTTGG 89452
Qy      330 ------------------------------------------------------------ 330
Db    89453 AAAGGGTGAATCAGGGAGAGGTTAAAGTGATTGGAAGGGGAGCGAGGGAGGGCGGTGGGG 89512
Qy      330 ------------------------------------------------------------ 330
Db    89513 GGCAGGCTTGGGGTACTGGGGGGGGGGCAGAGACCTTAGTTGCGGGATCCCGCGCTCAGC 89572
Qy      331 ----------------LeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerAr 345
Db    89573 GTCGCCCCTCGCGCCCCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCG 89632

Qy      345 gThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLe 365
Db    89633 GACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGAGCCGCGGCCAGCGCT 89692
Qy      365 uHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGl 385
Db    89693 GCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGG 89752
Qy      385 yGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVa 405
Db    89753 CGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGT 89812
Qy      405 lAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGl 425
Db    89813 GGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGA 89872
Qy      425 uGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLe 445
Db    89873 GGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTT 89932
Qy      445 uValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTy 465
Db    89933 GGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTA 89992
Qy      465 rGlnLysAlaArgGlyVal----------------------------------------- 471
Db    89993 CCAGAAGGCACGGGGTAC-GTCGGCCTGGGGCGGGATGCACCAGCGCAGGCCCGGGCGGG 90051
Qy      471 ------------------------------------------------------------ 471
Db    90052 GCGCAGCCAGAAGGGGTTCAGGTCGGATGACTGAAGGGTTGCCTGATGGGTACCTCTGCG 90111
Qy      471 ------------------------------------------------------------ 471
Db    90112 GCAAGGATGGACTGGGGTAGGAGCCCATAGGAGAGACCCGCATAAGTATGGTAGGTATGG 90171
Qy      471 ------------------------------------------------------------ 471
Db    90172 ATTAGGCAACTGGCGCTGGAAGGGCCACTGCTTCCCTCAGTCATGTTTGTTGAGAGAGGA 90231
Qy      471 ------------------------------------------------------------ 471
Db    90232 GGGAGAGGACGTTCCTGGAGCCACCTCCCATCCCAGTTAACTAACTGGAGAGGGGGGTCA 90291
Qy      471 ------------------------------------------------------------ 471
Db    90292 TTTCCAGAAAGCACTCCCTAACCAGTGGAGGAATTAATACACTGAGACAGGAAGCTGAGG 90351
Qy      471 ------------------------------------------------------------ 471
Db    90352 GAGTTTATACACTCCAGCTCAGATTATCTGGGATGATTTAGCTATAAATAGACACCAGGA 90411
Qy      471 ------------------------------------------------------------ 471
Db    90412 CAATATGAAGGATTCCTATACAGACTCCAGCTAGCCAGCACGGGGGGGTGGCACTATGAC 90471
Qy      472 -------------------------------AspAsnValGluTyrGlnPheAlaValAs 481
Db    90472 AGTCCCCTTGTCCCCCGTGTCACCAGGCATGGACAATGTGGAATACCAGTTTGCAGTGAA 90531
Qy      481 nAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTy 501
Db    90532 CAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTA 90591
Qy      501 rValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisTh 521
Db    90592 CGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACAC 90651
Qy      521 rLeuAspAspVal----------------------------------------------- 525
Db    90652 ACTGGATGATGG-TAGGGCCTCTGAACTCGCAGTGGGCAGCTTGGGGTAAAGAGAATTCC 90710
Qy      525 ------------------------------------------------------------ 525
```

```
Db  90711 CTGGGGAGGGGAGCTCACCCTCTCTTTACTAATGTAAAGGCATTTTTACCTGCTCATATC 90770
Qy    525 ------------------------------------------------------------ 525
Db  90771 ATCTTTTGTGCCCCGTAAGAGAGTGGGCGGGGTTCTTCTCCTCAGTTCACAGTTGGGGAA 90830
Qy    525 ------------------------------------------------------------ 525
Db  90831 ACTGAGGCCCAGAAAATGAGAATGATTTGCCCAGGTCAGTAGTTCAGCTGGGACTTAAGC 90890
Qy    525 ------------------------------------------------------------ 525
Db  90891 CCAGGTCTGTGATGATTCCCTGTGTCATGCAGTTCTCCCTGACACCTATGTGCCATCCAG 90950
Qy    525 ------------------------------------------------------------ 525
Db  90951 CTGGGGTAGGGGCCAGTGACATTAACCCATTAGACGTGGAAGATTGTGCTCCAAGACTTT 91010
Qy    525 ------------------------------------------------------------ 525
Db  91011 TTCCAGCTTCAGGGTGGGAAAGATGTCCCCTCATCCTTCAGCTTGGGGCTCCCCTGAGAA 91070
Qy    525 ------------------------------------------------------------ 525
Db  91071 GCTCCCTTTCCTCCCTCAGATCCTGGTCCTGGCCTACTTCTGACTTGCTGTGTGACCCAG 91130
Qy    525 ------------------------------------------------------------ 525
Db  91131 GCAAGTCTCTGACCGCTCCCGAGCCTCACCTTCCACGTCGGTGGAGCAGAGCTCATGCTG 91190
Qy    526 ------------------ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnPr 539
                            |||||||||||||||||||||||||:::||||||||||||||||
Db  91191 ATGTTCTTCTGTCTCAGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCC 91250
Qy    539 oSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLe 559
          ||||||||||||||||||||||||||||||||||    ||||||||||||||||||||::
Db  91251 TTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGT 91310
Qy    559 uLysTyrLysIleGluTyrGlyLeuGlyLysGlu------------------------ 570
          :||||||||||||||||||||||||||||||||||
Db  91311 GAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGGTGAGTGGGGGTGGGTGGGCTGCAG 91370
Qy    570 ------------------------------------------------------------ 570
Db  91371 GCATGTATGGGAATGATCAACGTGAAGAGCTGTCAGGGCTGCTGGAGACCTGAGCAGGGT 91430
Qy    570 ------------------------------------------------------------ 570
Db  91431 ACGGGAGTCGGAGGAAAGGAAGCAGAGTAGGGGTTGGGGAAGGGTGAACCTGAAAAACAG 91490
Qy    570 ------------------------------------------------------------ 570
Db  91491 GGGTCTGGGCAGGAATGAAAAGGGAAGACAGAGGGGCCTAGCTTAAGGCAGAGGGAGGCC 91550
Qy    570 ------------------------------------------------------------ 570
Db  91551 TGGGCTCTCACCCCAACCCCTCTACAGCTGCAGTGTGTACTTCTACTGAGTCCCTGCCTC 91610
Qy    570 ------------------------------------------------------------ 570
Db  91611 AGTTTCCTTATCTGGGAAATGGGGGGCTGAGGTGGAGGATCACTTGACCCCAGGAATTC 91670
Qy    570 ------------------------------------------------------------ 570
Db  91671 AAGGCTGCAGGGAACTGTGATCGTGCCATTGTACTCCCATCTGGGTAACAGAGCAAAACT 91730
Qy    570 ------------------------------------------------------------ 570
Db  91731 CTGTCTCAGAAAAAAATAAAAGGATAAAAAGTTAGGAGGAGGAAGGTAGCTGAGGGACCA 91790
Qy    570 ------------------------------------------------------------ 570

Db  91791 GGGGAAGGGTGAGGAAGGAGGGAAAGGTGAGGAAGGAGGGAAGGGTGAGGAAGTGGCCAT 91850
Qy    570 ------------------------------------------------------------ 570
Db  91851 GGTGGCCCTCAGGGCTCTGGAAGCCAGGCGTTCTTGCTGGCCTTCATGGTCAGCCTCTTG 91910
Qy    570 ------------------------------------------------------------ 570
Db  91911 TGTCTGCATCCCGTGTGTGTGTGTGTTGGCAGGGGATGGTCATTGACAGCTCCTGGGGAA 91970
Qy    570 ------------------------------------------------------------ 570
Db  91971 GTGGGGAAAGCCTGCCATCATTCTCCCTGCAGCTGAGCCCAGGCTGATGGCCTGCATCAG 92030
Qy    570 ------------------------------------------------------------ 570
Db  92031 CCAGAGTTGGAGGCTCAAGCGTCAACCCTGGAGCCATAGAATCTGCCCAGGCACTGAGGG 92090
Qy    570 ------------------------------------------------------------ 570
Db  92091 GTCAGTGCTGAGAACGCCTCAGGCTCCCACGGGCACGGAGGAACCGGGATGTGAGGCGGA 92150
Qy    570 ------------------------------------------------------------ 570
Db  92151 AGAGCCATGTAGGGGGGTTTAGGATCCTGTTCCTCAGGTGGCTACCCAGCAGCCCACCTT 92210
Qy    571 -------------AspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuTh 586
                       |||||:::||||||||||||||    ||||||||||||||||||
Db  92211 TTCTGTCCCTGCAGATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACACAGCTTAT 92270
Qy    586 rLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAl 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db  92271 GCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGTACAGCAGC 92330
Qy    606 aGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSe 626
           |||:::|||   ||||||||||||||    |||||||||||    :::||||||||||
Db  92331 CGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCACAACCAGAG 92390
Qy    626 rHis-------------------------------------------------------- 627
           ||||
Db  92391 CCATGGTACGGGGTTAGAGGAGAGACAGGACAGTGTGTGTGTGACTGTGTGGGTGTGCGT 92450
Qy    627 ------------------------------------------------------------ 627
Db  92451 GTGTCCTTGGGACCCAGCGCCCCACCCCCTGGCCCCGCCCTCCATCAGAGACCTTCCATG 92510
Qy    627 ------------------------------------------------------------ 627
Db  92511 GCTCCCTCCGCTTGATGGGTGAGTGAGGCTTAGGTGCAGAGGTGAAGGGCTGGGATCTGG 92570
Qy    627 ------------------------------------------------------------ 627
Db  92571 TGTCCCTTCGATTTGCTGTGTGACCTTGGACAAACCATTCTGAGGCTTCATTTACTGCCC 92630
Qy    627 ------------------------------------------------------------ 627
Db  92631 AGGAATGGCATGAGGTTGAATAATGTATTGTAACTAAAGTGCTTAGTACATAAAAACTCA 92690
Qy    627 ------------------------------------------------------------ 627
Db  92691 TTGAATTGTAGCTGCTTCGAGGATGATGGTGATGATGATTGTTTTTGTTGAGAGGCCCAA 92750
Qy    627 ------------------------------------------------------------ 627
Db  92751 TAGAAGAGTGATTAAATGCTGACCAAACCGGTCTACATGCTCTCTATGCGACACTAGGTA 92810
Qy    627 ------------------------------------------------------------ 627
Db  92811 AATTACTTAACTCTCTGAGCTTCAGTTTTCTTATCTGTAAAGCGAAGGTTGTAGTATTTT 92870
Qy    627 ------------------------------------------------------------ 627
Db  92871 CACTATTTTAGAGGGTTGCAAGGATAAAGTTATAATAACTACACAAAGTGCATAGGACAG 92930
```

```
Qy        627 ------------------------------------------------------------ 627
Db      92931 TGTTTGACACATTGTAAGCTCCTCACTCTCAAGTGGGTCTCCCTACTGTAATGGTGGCAT 92990
Qy        627 ------------------------------------------------------------ 627
Db      92991 CCCTCCTTCCTGGCCTCAGCAACCACCACCCCCACTGCCCCCAGTGTTTGATCCTTCAGA 93050
Qy        627 ------------------------------------------------------------ 627
Db      93051 AGACTTCACAGAGGCCGATTCTGTGCCAGGTGCTATCAGCCTACTCCAGAGGTGCTCCTT 93110
Qy        627 ------------------------------------------------------------ 627
Db      93111 GCTGACCCCTTTGCAAATGCCCAGCGGACACCCCTGCATTCCCTGCACTCTGTGCTTGTA 93170
Qy        627 ------------------------------------------------------------ 627
Db      93171 GAACAGTGAGTTCACCTCTAGACCAAGCCCAGGACTGGCACTTAGAGTGGTGGCTGAGTG 93230
Qy        627 ------------------------------------------------------------ 627
Db      93231 GTGATGTGTTGGGAGAATGCTGGCAAAAGGCTCTCACTGAATGCCATCTTCCTGTCTCTT 93290
Qy        627 ------------------------------------------------------------ 627
Db      93291 CCTCATGCCTCTTCCCTGTCTCCTATCCCTGGCCTCCTTCTCCTCATTCTGCATTCTGTT 93350
Qy        628 ------------------------------------ValProPheAlaProAlaGluLeuLy 636
                                                  ||||||||||||||||||||||||||
Db      93351 CCTTAATTTTCCCATCCTGCTCCTCTGCCCCCCAGTCCCTTTTGCCCCTGCAGAGTTGAA 93410
Qy        636 sValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGl 656
              ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93411 GGTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCA 93470
Qy        656 nIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAs 676
              ||||||||||||||||||||||||   |||||||||   ||||||||||||:::||||||
Db      93471 GATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGA 93530
Qy        676 pArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLy 696
              ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93531 TCGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAA 93590
Qy        696 sLysValLysGlnTyrGluLeuThrGlnLeu----------------------------- 706
              |||||||||||||||||||||||||||||||
Db      93591 GAAAGTGAAGCAGTATGAGCTGACCCAGCT-AGGTGAGGAGGGCTGGGGGGAGGGGGAGG 93649
Qy        706 ------------------------------------------------------------ 706
Db      93650 GAGGGATGTAGCTTCAGTGGGGACTGGGCAAGTTGTGTGAAGAAGTTAGAGATGGGATGC 93709
Qy        706 ------------------------------------------------------------ 706
Db      93710 TCTCTGTGGACGGGAGAGTGACCCCAGATGTCCCCCTAGCCATTAAGCCCTGTAACCTTC 93769
Qy        707 ----ValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGl 725
                  ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db      93770 CCCAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAGGATGG 93829
Qy        725 yTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAsp----------- 741
              ||||||||||||||||||||||||||||||||||||||||||     |||
Db      93830 CTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGG-TGAGGGCGGTG 93888
Qy        741 ------------------------------------------------------------ 741
Db      93889 GGGGAAGAAGGCGGGGAGGGCTCAGGTGAGGGCTCTGCTAGCCGCTTCCTCTTCTTTTCC 93948
Qy        741 ------------------------------------------------------------ 741
Db      93949 TTTGGGCCAGCCTGGTGAGCAGGTACTTTAGTAGCCTCAGCTCATGCCCAAACCCATTCT 94008

Qy        741 ------------------------------------------------------------ 741
Db      94009 CCTAGGATCAGGGGTCCTCTTTTAGCTCCAGAGAGCCTCTGTCATGGAGAGAAGCTTGTT 94068
Qy        741 ------------------------------------------------------------ 741
Db      94069 TGTTCACTGTCTTAGGCCTGCCACCTTATTTGCATTACCCATGAATAGTGCTGATTTCCA 94128
Qy        741 ------------------------------------------------------------ 741
Db      94129 AAGTTTGAACAGAGGGAAACCCTATGGTTGATCAGTAACTCCCGACCCCAAGTTTGACTG 94188
Qy        741 ------------------------------------------------------------ 741
Db      94189 TAAAACAAGTTTTTTAAAAACACATTTGTCTACATCACACTATCAGACTATCTCTCAAG 94248
Qy        741 ------------------------------------------------------------ 741
Db      94249 TCTAAGAAGAGTGAGTATCCCTGTGTGACCTAACGGGAAACTGAGGCTCAGAGAGGAAAA 94308
Qy        741 ------------------------------------------------------------ 741
Db      94309 GGAACTTGTCCAAGGTCACACAGAGCCTAGGCTACAACTGGGACTCCTGGTTTCCAGGCA 94368
Qy        741 ------------------------------------------------------------ 741
Db      94369 AAACTCTTTCCACTGCACTACACTGCCTTGTGGCCCTGCATTGGACACCTCAGTTGCCCA 94428
Qy        741 ------------------------------------------------------------ 741
Db      94429 TCATGATACTCCCTCACCCCTAACTTCCCAGCCACATGGTTTAGCAGAGGGAGTAATCAA 94488
Qy        741 ------------------------------------------------------------ 741
Db      94489 CCAGAAGCCATGAAGCTTGGGTTCTTTTTTTTTCTTTTGAGACAGAGTCTCACTCTGTTG 94548
Qy        741 ------------------------------------------------------------ 741
Db      94549 CCCAGGCTGGAGTGCAATGGTATGATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCA 94608
Qy        741 ------------------------------------------------------------ 741
Db      94609 AGCAATTCTTGTGCCTCGGCCTCTTGAGCAGCTGGGATTACAGGTGTGCACCATACCTGG 94668
Qy        741 ------------------------------------------------------------ 741
Db      94669 CTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGCCTTGAA 94728
Qy        741 ------------------------------------------------------------ 741
Db      94729 CTCCCGACCTCAGGTGATCCACCCATCTTGGCCTCCCAACGTGCTGGGATTACAGGTGTG 94788
Qy        741 ------------------------------------------------------------ 741
Db      94789 AGCCACCACACCCAGCTGAGTTCTAATTCCAGTTTGTTCATTTCTTCATTCAGCAAATAT 94848
Qy        741 ------------------------------------------------------------ 741
Db      94849 TTCGCTGAGTGTCTACTAAGTGCCAGCTGCTTTGGGAATTTAGTGGTGGACAACACAAAT 94908
Qy        741 ------------------------------------------------------------ 741
Db      94909 AAGGTTCTTGCTCTCTTGGGGTTTATTTTATGTTTCTAAAATAAACACTTATAAATATAG 94968
Qy        741 ------------------------------------------------------------ 741
Db      94969 TGTTTATTATATATTCAATTTCCTATCTATTTTTAAAATATTAAATAATTTAATCCTCATA 95028
Qy        741 ------------------------------------------------------------ 741
Db      95029 ACAACCCCATGAGGTAGGTACTATCATTAGAATAGTGAATGGAGACACAGAGAAGGTAAG 95088
Qy        741 ------------------------------------------------------------ 741
```

```
Db      95089 TAATTTCTCCAAGGTCACACAGCTAGTAAGTGTCAGAGCTGGAATTTGAACCAAGGCAGC 95148

Qy        741 ------------------------------------------------------------ 741

Db      95149 CTGGTTCCAGAGTCTGGTTCCATGGTTATTCATTGCTGTGATTTGCTGCCTCTCTCTTTA 95208

Qy        741 ------------------------------------------------------------ 741

Db      95209 TTACGTTCTTACAGACTAGTGTTCCACAGTCAAAAATCAAATTCTTGGAAGCCTCAATTG 95268

Qy        741 ------------------------------------------------------------ 741

Db      95269 CCTTATCTGTGAAGTAGGAATGATGCCACTCAACTCACAGGATCAAAGGAGTGGGTCAAA 95328

Qy        741 ------------------------------------------------------------ 741

Db      95329 GAAACAGCCTTGGGAACTGTAAACCACTATACATATTTAAGGATGAGTATCCTCCCTCCC 95388

Qy        742 -------------------------------LeuProIleGlnArgGlyProProLeuPr 751
                                             :::|||||||||||||||||||||||||
Db      95389 TTCCTCACACTCTAGTTTTCACCTTTCAGACATGCCTATCCAGAGGGGACCACCCCTGCC 95448

Qy        751 oProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLy 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95449 TCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAA 95508

Qy        771 sProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLe 791
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db      95509 GCCAGATTTCACCACAGTCAAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCT 95568

Qy        791 uArgAsnAlaSerLeuValThrTyrTyrThrSer-------------------------- 802
              ||||||||||||||||||||||||||||||||||
Db      95569 CAGGAATGCCTCCCTGGTCACCTATTACACCAG-GTGGGGACTAATGCCTTTACTAATGT 95627

Qy        802 ------------------------------------------------------------ 802

Db      95628 AAAGAAGGGAGGAAGAATAGGTGGGGGGTGGTGGTGGGAAGGCAGCAGGTGGTGATCAGG 95687

Qy        802 ------------------------------------------------------------ 802

Db      95688 GAAGGGCCCCTGGGATTCAGGAACATGTCATGGCTCACCTGTTACGCTGAGGCCACACAG 95747

Qy        802 ------------------------------------------------------------ 802

Db      95748 AGGGGCTGCACAGGGTCCACGTCCTTGCCTGCTGACCAGATGGTAGCAAGGAGGCCATCC 95807

Qy        802 ------------------------------------------------------------ 802

Db      95808 AATCCCGTCCCCTCACAACAGAGCCTTGTTCTCAAACTCCGGAGAATGTGACACAGGACA 95867

Qy        802 ------------------------------------------------------------ 802

Db      95868 GACCTGTTTCTCCAACCTCAGAGGGCTCCCAGTCTGATGGTGAAGAGAACCCCACCCCTG 95927

Qy        802 ------------------------------------------------------------ 802

Db      95928 TCCTCAGAGCTGCTCCCTCTACCCTGTTAGATGGGGAGACAGAGCCTGAGGAGCCTGAGG 95987

Qy        802 ------------------------------------------------------------ 802

Db      95988 TTTGGCGGAGGAAAGTCGCATAAAAAGCAACCGATATATACAACTGCCATGAAGGCAAAA 96047

Qy        802 ------------------------------------------------------------ 802

Db      96048 TTAAGCCAAAGAAAAGATGCATCGGAGATGATTCCATAGGTGGAGAATAACTTGGATTAA 96107

Qy        802 ------------------------------------------------------------ 802

Db      96108 TCCAGGAAGGTTTCCTGGAGGAAGGGGTCTGGTTTCGAGATTGGACTTGGAGCAGCATAC 96167

Qy        802 ------------------------------------------------------------ 802

Db      96168 ATCCAGACTTCTCTGTAGGTCCCTGGGTGGTGGTAGTCGCCTCCAGGCCAGGGAGCAGCT 96227

Qy        802 ------------------------------------------------------------ 802

Db      96228 TCCCTTCCCAGGGTAGAGCCTGGGGCTGGGGCTGACTGCAGGGGAGTAGATTTGGAGGAG 96287

Qy        803 ----------------------------------------------------------Se 803
                                                                        ||
Db      96288 TCTGCCCTGCTGGGAGAGTGTGCCTGAGAGCACTAGTCACGTTTCTTGCCTTAGCAGTTC 96347

Qy        803 rGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGl 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96348 TGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCA 96407

Qy        823 nSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThrLeuPr 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96408 GTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCC 96467

Qy        843 oAspArg----------------------------------------------------- 845
              |||||||
Db      96468 TGACCGTGAGTGGCTCCCAGCCCCTTGCCACCAACCCATCTTTGCCCAACTCCTTCTTG 96527

Qy        845 ------------------------------------------------------------ 845

Db      96528 CCCCAGCAGCAAGAAACATACTTCAGAACCCCAAGGTTTGGGCAGAGCTGGGGGCAGAAG 96587

Qy        845 ------------------------------------------------------------ 845

Db      96588 AGTTTCTGGGAGGGGGAGGAATGTTAAGTCCCTCCAACCACACATACTGTTTTCTAGAATT 96647

Qy        845 ------------------------------------------------------------ 845

Db      96648 CCCCAACTAGAAACCAGACCTCATCTTTGCTTTGTGGCTTGTTCCATGAAGGAGGAAGAA 96707

Qy        845 ------------------------------------------------------------ 845

Db      96708 GGGAGGTCATATCATCATGCTCTGTCCCCCAGCTCTGACATCTCTCCCCATCTCCTCTCC 96767

Qy        846 -------------ProSerThrProProSerAspLeuArgLeuSerProLeuThrProSe 861
                           |||    ||||||||||||||||||||||||||||||||||||||
Db      96768 CGCACCCCAGGGCCCTCC-ACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTC 96826

Qy        861 rThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLe 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96827 CACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCT 96886

Qy        881 uIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGl 901
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db      96887 GATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCA 96946

Qy        901 u----------------------------------------------------------- 901

Db      96947 -GGGTGAGGGCTCCTGCCCCAGGGCATCTGCTCAAGAGCTTCCCAGCCTCTCCCGCTTCC 97005

Qy        901 ------------------------------------------------------------ 901

Db      97006 AGCAGTGAGAACAGTGCTGGGGTTTGTTCCAGGGGCTACTGGAGAAGCAAGCAGAAGCT 97065

Qy        901 ------------------------------------------------------------ 901

Db      97066 GCTGCCTAAGTGGAGAGCAGACTCTCCCAGTCCCCTCCCTGCCCACCTCTGCATCCACCC 97125

Qy        902 -------------------------------------------GlyAsnIlePheSerAl 907
                                                         |||||||||||||||||
Db      97126 ACCCCATCCACCATCCCTGCTCTCCCTGCCTCCCTCACCCCCAGGAAACATCTTCAGTGC 97185

Qy        907 aGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGl 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97186 TGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGA 97245

Qy        927 uValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSe 947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db      97246 GGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTC 97305
```

```
Qy        947 r----------------------------------------------------------- 947
              |
Db      97306 AGGTATGAGCACAGGGCAGCCCAGGTTTCCCTGGATTGACTTCCAAGCCCCTGCCTGTGG 97365

Qy        947 ------------------------------------------------------------ 947

Db      97366 ACACAACTACAATTTCATTGTTTTGCACAGCTGTCCTCCAGGTTTGGTGGGGGCAGAGCA 97425

Qy        947 ------------------------------------------------------------ 947

Db      97426 AATGTGAAGGGGAAATAGGTGTGGTTCAGTGATAACTTGATTCAGAGGCCAGGTTCGTGG 97485

Qy        947 ------------------------------------------------------------ 947

Db      97486 GGGACTGGCTTTTTTTTTTTTTTTCTGAGATGGAGTTTCACTCTTGTTGCCCAGGCTAGA 97545

Qy        947 ------------------------------------------------------------ 947

Db      97546 GTGCAATGGCGCAATCTCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCGATTCTGC 97605

Qy        947 ------------------------------------------------------------ 947

Db      97606 TGCCTCAGCCTCCCAAATAGCTGGGATTACAGGCACACACCACCATGCCCGGCTAATTTT 97665

Qy        947 ------------------------------------------------------------ 947

Db      97666 TGTATTTTTAGTAGAGACAGGATTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 97725

Qy        947 ------------------------------------------------------------ 947

Db      97726 CTCGTGATCTGCCTACCTTGGCCCCCTAAAGTGCTGGGATTACAGGCGTGAGCCACCATG 97785

Qy        947 ------------------------------------------------------------ 947

Db      97786 CCCAGCCAATTATTGTATTTTTTAGTAAAGACGGGTTTTCACCACATTGGCCAGGCTGGT 97845

Qy        947 ------------------------------------------------------------ 947

Db      97846 CTCAAACTCCTTACCTCAGGTGATCCACCCACCTCGTCCTCCCAAAGTGCTGGGATTACA 97905

Qy        947 ------------------------------------------------------------ 947

Db      97906 GGTGTGAGCCACTGCATCCAGCTGGGACTGGCTTTAGAGACCAGTTTCTTTGGGAAATTG 97965

Qy        947 ------------------------------------------------------------ 947

Db      97966 CTTTGATTGGAGGATTAACTTGGTGAGGTGATGAGCTTTATGGTGTGGAGGTGGACAGCA 98025

Qy        947 ------------------------------------------------------------ 947

Db      98026 GGACTGGCTACGTAATTTCCAGGGCCCAGTGCAGAATGAAAAGGTGCTATCCCTTATTTT 98085

Qy        947 ------------------------------------------------------------ 947

Db      98086 AAAAATTAGAGGTCGGACGGCCAGGCGCGGTGGCTTATGCCTGTAATCCCAGCACTTTGG 98145

Qy        947 ------------------------------------------------------------ 947

Db      98146 GAGGCCAAGGTGCGGGCGGATCATGATGTCAGGAGATCGAGACCATCCTGGCTAACACGG 98205

Qy        947 ------------------------------------------------------------ 947

Db      98206 TGAAACTCTGTCTCTACTAAAAATACAAAAAAAATTAGCTGGGCGTGGTGGTGGATGCC 98265

Qy        947 ------------------------------------------------------------ 947

Db      98266 TGTAGTCCCAACTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGAGGAG 98325

Qy        947 ------------------------------------------------------------ 947

Db      98326 CTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCCG 98385

Qy        947 ------------------------------------------------------------ 947

Db      98386 TCTCAAAAACAAAAACAAAAACAAAAAAAGCTAGAGGTCGGGCATGGTGGCTTGTGCTTG 98445

Qy        947 ------------------------------------------------------------ 947

Db      98446 TAATCCAAAGTGCTTATGCACTTTGGAAGGCTGAGGTGAGAGGATCACTTGAGGCCAGGA 98505

Qy        947 ------------------------------------------------------------ 947

Db      98506 GTTGGAGACCAGCCTGGACAACATGGTGAGACCCTGTCTCCACAAAAAATTTAAAAATTA 98565

Qy        947 ------------------------------------------------------------ 947

Db      98566 TCTGGGTGTGGTAGCTCATGCCTGTGGTCACAGCTGCTTGGGAGGCTGAGGTGGGAGGAT 98625

Qy        947 ------------------------------------------------------------ 947

Db      98626 CACTTGAGCCCAGGAGGTTGATGCTACAGCGAGCCGTGTTCATGCCACTGCACTCCGGCC 98685

Qy        947 ------------------------------------------------------------ 947

Db      98686 TGGGTGACAGAGCAAGACCCTGTCTCAAAAGAAAAAAAATTAGAATTTTAAGGTGGTGAC 98745

Qy        947 ------------------------------------------------------------ 947

Db      98746 AGCAGAGCATGAACTCAAGCGTGGGGTCCATCCAAGCATGGGGTCTGCGCACCTGCACAT 98805

Qy        947 ------------------------------------------------------------ 947

Db      98806 GTTGCACACCCATGCAACAGCCCTGGGAGCAGGCAACACGTGCTTCACCAAAGGCCAGTT 98865

Qy        947 ------------------------------------------------------------ 947

Db      98866 TGAGGCAGAGAGACCATAACTTCTGGGTGTGTTTTAGGGAGGACAGTCAGAGGGAGAGGT 98925

Qy        947 ------------------------------------------------------------ 947

Db      98926 AGTAGTAGTTTCTGTGAGGCTCAGGGCCCTCCTGTCCTTGTATCAGAAGATTGTGAGCAT 98985

Qy        947 ------------------------------------------------------------ 947

Db      98986 CTGAAGGGTGAGGGTCCAGCTGGTGTCTCCTGAGAAAGCCATGGATAGTCCTGATCTATG 99045

Qy        948 ---------------------------------------------------AspSerL 950
                                                                 |||||||
Db      99046 CCCTGTGCCCTTCCTCTGTCCTGACCTGCTCTGTGCCTCTCTCTCCCCCACCAGACTCGC 99105

Qy        950 euAspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuL 970
              ||||:::|||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db      99106 TGGACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCC 99165

Qy        970 euAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArg------------------- 983
              |||||||||||||||||||||||||||:::|||   |||
Db      99166 TGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAG-GTGAGCAGGGGAGCCAGTC 99224

Qy        983 ------------------------------------------------------------ 983

Db      99225 TGACTTCAGGGAAGGGAGACTTCCAGGAAGGGCTGCAGAATGGCCATCAGGGAAAGTGG 99284

Qy        983 ------------------------------------------------------------ 983

Db      99285 CCATGACTTGCCCTTGGCACTCCCAGGAGCCCCGTTAAAGTCAGGACTCTGGGATGAAGG 99344

Qy        983 ------------------------------------------------------------ 983

Db      99345 AATCAGGGATGATGTCCTCCTGAGAATGGGGAGTGGACTTAGGGGAGCCCCAGCCTACCC 99404

Qy        983 ------------------------------------------------------------ 983

Db      99405 TCCTCAGCAAGGCCAACAGAACTGCCTCACTCCCACTTCCTCTGTAGCTGCCAGCATCTT 99464

Qy        983 ------------------------------------------------------------ 983
```

```
Db   99465 GGAAAGAGGGGCCCCAAATCTCCAGATGCCTGCAGCCAAGGGAAGAAGAGCCCTGTTTAC 99524
Qy     983 ------------------------------------------------------------ 983
Db   99525 AGCATCAGGCTGTTTTGAACAATAGCAGTTTTTCTTACTCATTGCACAGACCAGGAAAGC 99584
Qy     983 ------------------------------------------------------------ 983
Db   99585 AAATCCCAGAGATGGGCAGGAACCTCCCCAGAGACATAGGCTGGACCAGATCCCAGCCCC 99644
Qy     984 -----------------------------------------------------GluAlaL 986
Db   99645 GCCTGACCCCTTACTCTGACCGCTATGCCTTGTTTTCCGTCTTCCCTTCCAGGGAATCCC 99704
Qy     986 euProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgL 1006
Db   99705 TCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGAGCTCGGC 99764
Qy    1006 euGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisProArgProG 1026
Db   99765 TTGGCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCCCATCCCC 99824
Qy    1026 lnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSerLeuMetG 1046
Db   99825 AGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGCCTTATGG 99884
Qy    1046 lyGlySerValSerAspCysArgGlyHisSerLysArgLys------------------- 1059
Db   99885 GTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGGTAAGTCAGAGCCCCACAT 99944
Qy    1059 ------------------------------------------------------------ 1059
Db   99945 TGCGGCTCTTTTCTAAATAGGACACCAGGGGGCGACAGAGCACAACGCGTGGCTATGCGA 100004
Qy    1059 ------------------------------------------------------------ 1059
Db  100005 GCATGGCTCTACCTCCTCTTCGGAGCCGGCTGCGGACGCTAGGGCCTCTCTCCCTCCTCC 100064
Qy    1059 ------------------------------------------------------------ 1059
Db  100065 ACGGAATGGTTGGCTGTCAGGGAAGCACATGGGCCTCCATATCCCGGCGGAATAGGTCAG 100124
Qy    1059 ------------------------------------------------------------ 1059
Db  100125 CTTGGTTCTCACGAAGGAAAGGGCCTCCCAGCAGGTGACCCAGGGTAGCCAGAACACCCA 100184
Qy    1059 ------------------------------------------------------------ 1059
Db  100185 GCTCCCTCCTCATCCCTGGAGCTGGGGAGACCTCTGAGACCCAGACATCCCCGTGTGGGG 100244
Qy    1059 ------------------------------------------------------------ 1059
Db  100245 AGAAGAGGCCAGTCGGCTTGGCTGTGGCCAAGTGGGACTGGACCAGGATAAGGTGAGGC 100304
Qy    1059 ------------------------------------------------------------ 1059
Db  100305 CTTCGCTGGGCTGGTCCCTTTCATGCTGGCAGGGGCGGCACATCCCACCCAACTCCGACC 100364
Qy    1059 ------------------------------------------------------------ 1059
Db  100365 CGTCCTCCCTCACCCTCTCCCCGCCTGGCTCGGACACCGCGGGCTTTGCAGAGCGTCTGG 100424
Qy    1059 ------------------------------------------------------------ 1059
Db  100425 TGTTCCGCTCCTCTGCAGGAGGACTCATCCTCAGATTACCCCGCTGACTCTGAGCCCTTT 100484
Qy    1059 ------------------------------------------------------------ 1059
Db  100485 CGATGCTCCTCCCCAGAGCCTTGGAGAGGGAGAGAGCAGCGGGTGACAGTCTGGACCCTG 100544
Qy    1059 ------------------------------------------------------------ 1059

Db  100545 CTCCCCCGCGGCCCATCCCAATTCCCTCCCCCGGGCTTCCCTCTCCCAGCCAATCCTGGT 100604
Qy    1060 ------------------------------------IleSerTrpAlaGlnAlaGlyGly 1067
Db  100605 CGTGCCCCTTACTGTCCCCTTCTTTCCCCCGTGCAGATCTCCTGGGCTCAACCAAGCGGG 100664
Qy    1068 ProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerGlyProArgPro 1087
Db  100665 CTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAG---GCAGGCCCCCGGCCG 100721
Qy    1088 AlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAla 1107
Db  100722 GCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGCT 100781
Qy    1108 LeuValTyrAspGlyIleLys--------------------------------------- 1114
Db  100782 CTGGTGTACGACGCCATAAAGGTCAGCGTCCTGGAAGGAAGGGGCAGGAAAGGAGGAAAG 100841
Qy    1114 ------------------------------------------------------------ 1114
Db  100842 GGTCTCTGTGCTTGTGAAGAGGAAGTAATGAGTTCTCTGGAGACCCCTGGGTGGTAACTT 100901
Qy    1114 ------------------------------------------------------------ 1114
Db  100902 TCCTGGGGATAGAAGTCTAGCCAGGAGACCCCTGGAGCCTCCAAAGGTTGAGTAGGTGGA 100961
Qy    1114 ------------------------------------------------------------ 1114
Db  100962 GGGCAGGCCAGTTTGCAGGAACCTGTTTCAACCTGGGTTGAAAACAGCCCCTCCTCTACC 101021
Qy    1114 ------------------------------------------------------------ 1114
Db  101022 CAGACCTAAGTGGGCCACAGCAGTTCCGTGGATCTGAGGGTGTGGCTGGCCTGAGTACGC 101081
Qy    1114 ------------------------------------------------------------ 1114
Db  101082 TCAGCCCTTCCTCCTGTCCCCAGAGGCCCTGACTGAGTCTGGTACTCCTGAAACTCCTGG 101141
Qy    1114 ------------------------------------------------------------ 1114
Db  101142 ATGAGGGCAAATGGGAAGTATTCCATCATTCCCTGCCTTTGTATAGGTTCGAACAGTTCC 101201
Qy    1114 ------------------------------------------------------------ 1114
Db  101202 TGTGTATCCTTCACCGACTCCCATGTGATACTTACTCTGATGGAGGGTCAGACTTTCTGC 101261
Qy    1114 ------------------------------------------------------------ 1114
Db  101262 CCCACGTACGCTGGTCATTCTGATGATTACTCTCAGTAATGGTCACTTTGTCTTTGTGTT 101321
Qy    1115 ---------------SerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnVal 1129
Db  101322 CTCTGCTCTGATCAGGGCAATGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTG 101381
Qy    1130 GluAlaGluValIleValHisSerAspPheGlyAlaSerLysGlyCysProAspLeuHis 1149
Db  101382 GAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCAT 101441
Qy    1150 LeuGlnAspLeuGluProGluGluProLeuThrAlaGluThrLeuProSerThrSerGly 1169
Db  101442 CTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGT 101501
Qy    1170 AlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnPro 1189
Db  101502 GTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTG 101561
Qy    1190 ThrThrSerGlyProGluArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSer 1209
Db  101562 GCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCC 101621
Qy    1210 CysSerAspLeuGlnProSerThrAlaIleGluGluAlaProGlyLysSerCysGlnPro 1229
Db  101622 TACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTC 101681
```

```
Qy       1230 LysAlaLeuCysProLeuThrValSerProSerLeuProArgAlaProValSerSerAla 1249
              |||:::  ||||||||  |     ||||||   ||||||||:::|||||||||||:::
Db     101682 AAATCCCCTGCCCTCTAGGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 101741

RESULT 15
AC112162/c
LOCUS       AC112162              116580 bp    DNA     linear   ROD 21-SEP-2002
DEFINITION  Mus Musculus chromosome 9 BAC clone MGS1-296M6 ES cell line,
            complete sequence.
ACCESSION   AC112162
VERSION     AC112162.8  GI:21327375
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   3  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   4  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   5  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   6  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On Jun 6, 2002 this sequence version replaced gi:21321767.
            ----------------Genome Center:
            Center:   Harvard Partners Genome Center
            Center Code:   HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact:  hpgc@mendel.mgh.harvard.edu

            CLONE LENGTH: This sequence represents the entire insert of this
            clone unless otherwise noted. If there are overlapping clones, the
            overlaps are noted in the beginning and end of the Features
            listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550).
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
              Genes and Regions of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
            and cDNA sequences in Unigene. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintain sequence continuity
            across the splice junctions. Sequences that are not identical
            matches are annotated as similar.

            SEQUENCING READ COVERAGE: Attempts are made to complete double
            strandedsequence for all regions.  All sequence is completed to a
            standard of coverage with a minimum of 3 reads with no ambiguities.
            If the sequence coverage for a region does not meet this standard,
            it is indicated in the annotation asLow Coverage.  Low coverage
            linkages are verified by PCR product size verification or
            verification of forward and reverse reads from clones which span
            the low coverage area.

            QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
            standards - estimated average error rate is less than 1 per 10,000
            bases using the Consed quality parameters.  Regions that do not
            meet this requirement are annotated as Low Quality.

            ----------------Summary Statistics
            Center project name: ALI
            Sequencing vector: pSMART; AF399742
            Chemistry: Dye-terminator Big Dye; 100%
            Assembly program: Phrap version 0.990319
            Contig length: 116580
            Fraction of Phrap value < 40: 0
            Error Rate in Consed: 0.00 per 10,000 bases
            Number of N's in consensus: 0

            ------------ Distribution of Quality < 40 Bases:

                    1000|
                     900|
                     800|
               #     700|
             bases   600|
                     500|
                     400|
                     300|
                     200|
                     100|
                       0|
                         -----------------------------------------
                           5    10    15    20    25    30    35    40
                                        Phrap Value Range
            ----------------------.
FEATURES             Location/Qualifiers
     source          1..116580
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="ES cell line"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /clone="MGS1-296M6"
     STS             235..369
                     /standard_name="D9Mit316"
     repeat_region   258..303
                     /rpt_family="(TG)n"
     STS             612..1204
                     /standard_name="Nope"
     repeat_region   complement(1584..1643)
                     /rpt_family="B4A"
     repeat_region   1710..1748
                     /rpt_family="(TG)n"
     repeat_region   complement(3793..3909)
```

```
                     /rpt_family="PB1D10"
     repeat_region   4042. .4191
                     /rpt_family="MIR"
     repeat_region   4247. .4435
                     /rpt_family="B2_Mm2"
     repeat_region   complement(4504. .4623)
                     /rpt_family="RSINE1"
     repeat_region   4720. .4831
                     /rpt_family="L2"
     repeat_region   complement(4840. .4879)
                     /rpt_family="MIR"
     repeat_region   4938. .5118
                     /rpt_family="B2_Mm2"
     repeat_region   6216. .6250
                     /rpt_family="AT_rich"
     repeat_region   complement(6251. .6381)
                     /rpt_family="RSINE1"
     repeat_region   complement(6421. .6566)
                     /rpt_family="B1_MM"
     repeat_region   6593. .6612
                     /rpt_family="(T)n"
     repeat_region   complement(6668. .6790)
                     /rpt_family="B1_MM"
     repeat_region   complement(6705. .6820)
                     /rpt_family="MT2B"
     repeat_region   complement(6883. .6954)
                     /rpt_family="ID_B1"
     repeat_region   complement(6982. .7058)
                     /rpt_family="ID4"
     repeat_region   8050. .8246
                     /rpt_family="B2_Mm2"
     repeat_region   complement(8445. .8719)
                     /rpt_family="B4"
     repeat_region   complement(9019. .9119)
                     /rpt_family="MIR"
     repeat_region   9265. .9371
                     /rpt_family="MIR"
     repeat_region   complement(10773. .10846)
                     /rpt_family="PB1D10"
     repeat_region   10859. .10956
                     /rpt_family="B4"
     repeat_region   complement(11043. .11141)
                     /rpt_family="PB1D7"
     repeat_region   11848. .11884
                     /rpt_family="(TC)n"
     repeat_region   11884. .11922
                     /rpt_family="(TG)n"
     repeat_region   11930. .11966
                     /rpt_family="(TC)n"
     repeat_region   complement(11967. .11994)
                     /rpt_family="B4"
     repeat_region   complement(11995. .12127)
                     /rpt_family="B1_MM"
     repeat_region   complement(12128. .12143)
                     /rpt_family="B4"
     repeat_region   complement(12170. .12252)
                     /rpt_family="RSINE1"
     repeat_region   12853. .12993
                     /rpt_family="B1_MM"
     repeat_region   complement(13288. .13367)
                     /rpt_family="MIR"
Alignment Scores:
Pred. No.:              2.84e-90        Length:          116580
Score:                  2466.50         Matches:         837
Percent Similarity:     20.68%          Conservative:    19
Best Local Similarity:  20.22%          Mismatches:      41
Query Match:            37.35%          Indels:          3248
DB:                     10              Gaps:            20

US-09-754-997A-2 (1-1252) x AC112162 (1-116580)

Qy        233 GlySerLeuGluAlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThr 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12486 GGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACC 12427

Qy        253 ThrValValSerGlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrPro 272
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db      12426 ACGGTAGTGTCTGGACAGAGTGTAGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCT 12367

Qy        273 PheValSerTrpValArgGln--------------------------------------- 279
              |||||||||||||||||||||
Db      12366 TTTGTGTCCTGGGTCCGACAGGGTGAGTGCTGAGGAAAGAGGAGGAGGAGACTGGGGAAG 12307

Qy        279 ------------------------------------------------------------ 279

Db      12306 TGGGCAAGATGGTACAGTATTTCACAGGCTCCAAGTCAGGAGCTTCTGCCTAAAGATAGA 12247

Qy        279 ------------------------------------------------------------ 279

Db      12246 AGGAGAGAATTGACTCCTTAGAGTTGTCCTTTGACTTCCATAGCACATGCAAGCAAGCTC 12187

Qy        279 ------------------------------------------------------------ 279

Db      12186 ACATGTACACACATTCATGCAGTAATAAAGTCTAAAGAAGGTTGTATTAAAGCCCAGTGG 12127

Qy        279 ------------------------------------------------------------ 279

Db      12126 TGGTGGCACACATCCTTAATCCCAGCACTGGGGAGGCAGAGGCAGGCGGGGCTCTGAGTT 12067

Qy        279 ------------------------------------------------------------ 279

Db      12066 TGAGGCCAGCCTGGTCTACAGAGCAAGTTCAGGGTAACCAGGGCTACTCAGAGGAAACCC 12007

Qy        279 ------------------------------------------------------------ 279

Db      12006 TACTTTGGAAAATTCTGGGGGTGGGAGGAGAGAGAGACACAGAGAGACAGAGAGAGCGAG 11947

Qy        279 ------------------------------------------------------------ 279

Db      11946 GGACAGAGAGACAGAGACAGGCAGACACACACACACACACACACACACACACACACACAC 11887

Qy        279 ------------------------------------------------------------ 279

Db      11886 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTGTATTAAATATCTGCTCCG 11827

Qy        279 ------------------------------------------------------------ 279

Db      11826 CCCACCAGAGACCATTTAGCCATCGTTTACCCACCTCCCTTCTGCTTTTCCACCTAGGGA 11767

Qy        279 ------------------------------------------------------------ 279

Db      11766 TCAGAGTAAGAAGGAACTGGGATTCAAACCAGCGGCTGTGATGCAGCTTTGCAGGGATGT 11707

Qy        279 ------------------------------------------------------------ 279

Db      11706 AGTGGGAACCTCGTAGTTGGAGAGGGTGTAGAGGGACAGTTCCAGCGTCCCCCTCCAGCC 11647

Qy        280 ------------------AspGlyLysProIleSerThrAspValIleValLeuGlyAr 293
                                ||||||||||||||||||||||||||||||||||||||||||
Db      11646 TGCTCCTTCTCTGTCACCAGATGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCG 11587

Qy        293 gThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAl 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11586 GACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGC 11527

Qy        313 aAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeu----- 331
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db      11526 CAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGGTAA 11467

Qy        331 ------------------------------------------------------------ 331

Db      11466 GGAAGGGCTGGTGCTGGGGGACGTTAAGGATGGGTGTGTGTTAGAGGGAGAAGGGGCGGT 11407

Qy        331 ------------------------------------------------------------ 331
```

```
Db    11406 CCTAGAAGACTAAATGTGAGGGGGACCACATGGAGGAGTCTGAATTTTCTGATTGGTGG 11347
Qy      331 ------------------------------------------------------------ 331
Db    11346 TGATTGTTGGAGGACTGGATTCCGTAGCTGGGGGGGGGGGGGGAGGGAATGGGGCGATTT 11287
Qy      331 ------------------------------------------------------------ 331
Db    11286 GAGACGGGGGGTGGGGGGGCGGATTTAAAGCCTGGTTTGGGTCATGAGGAGTGAGAGTGC 11227
Qy      331 ------------------------------------------------------------ 331
Db    11226 AGGCGAGCAGGAGTGGGGTGAGGTTATGGAAGGAGATAGTGGAATGGTCTAGGGGGGTGT 11167
Qy      331 ------------------------------------------------------------ 331
Db    11166 GTGAAGCCCAGAATCTGTACGTGGTAGCACTCAGAAAACTGTGGCAGGAGCGTGAGTTCG 11107
Qy      331 ------------------------------------------------------------ 331
Db    11106 ATACCAGCCCTAGCTACAGAGTTGAGACCAGCTACAGAGTGAAAAGTCTATCTCAAAAAC 11047
Qy      331 ------------------------------------------------------------ 331
Db    11046 AAAATTATTAAACAATGACAATGTCTTAGGCTATTAAGTCCCTATGAGTGTGTGCAGGCC 10987
Qy      331 ------------------------------------------------------------ 331
Db    10986 ATGGTGAGTGTATTGCGGTCACAGACTGAGTTGAGACAAGGTTTCTTGTTCACTTCATAC 10927
Qy      331 ------------------------------------------------------------ 331
Db    10926 TCCAGGCTAGCTGGCCCGGGGAATTCTGGGGAATCTCCTGTCTCTGTCTCCTAGGAATGC 10867
Qy      331 ------------------------------------------------------------ 331
Db    10866 TGGGATTAATCAGTTGGGGTGATGCTGAGGCGGGCAGATCAGAAATTCAAGGCCAACCTG 10807
Qy      331 ------------------------------------------------------------ 331
Db    10806 GGCTATAAAAGAGCTTTCTCAATACAACAGAAAACTGGGGGACCAGGCTTGCTGAAAGCT 10747
Qy      331 ------------------------------------------------------------ 331
Db    10746 GGAGTAACTGCTGGGTATAAGAAGCCACGGGGAGGAGAAAGTGAAGTTACAGCCTAGAGC 10687
Qy      331 ------------------------------------------------------------ 331
Db    10686 TCAAAAGCTGGAAATCTGCTGGGCTTGGGAGAAAGGAGGACTAGCCCAGGAGGGGTCGAG 10627
Qy      331 ------------------------------------------------------------ 331
Db    10626 TCTAGTGGACCAGAGCTGGGCCCGTATGACTAGGGGAGAGGACCTCAGGGTGAGGCTTGG 10567
Qy      331 ------------------------------------------------------------ 331
Db    10566 GCGCGGTGTTGGGACACTTGGAGAGGGGTAGGCCCTCTTGTCAGAGTCTCCCTTCTGCCC 10507
Qy      332 ----AlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerTh 350
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10506 CTCAGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCAC 10447
Qy      350 rAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAs 370
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10446 CGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGA 10387
Qy      370 pGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuVa 390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10386 CGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGT 10327
Qy      390 lIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAl 410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10326 CATCACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGC 10267

Qy      410 aGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGlyLeuProSerAl 430
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10266 GGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGC 10207
Qy      430 aProThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAlaTrpGluAr 450
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10206 CCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCG 10147
Qy      450 gProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGl 470
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10146 GCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGG 10087
Qy      470 yVal-------------------------------------------------------- 471
            |
Db    10086 TAT-GTCTACTCTTGACCGCGGGAATCCCAGTTAGGGCAAGTATGGGATGGATAATTAAA 10028
Qy      471 ------------------------------------------------------------ 471
Db    10027 GATGAGAGGAAGGGATGGGGAGCAATACTGCTGGGAAGTGATCTAGTATTAGAGAATGGA 9968
Qy      471 ------------------------------------------------------------ 471
Db     9967 GTGTACACGCACTCCTTGACGGAGAGAATGCTTCTCTTTCCAGGGTTAAAGGAGCCATCT 9908
Qy      471 ------------------------------------------------------------ 471
Db     9907 TCCACCCCCAGCTAACTAACTGCAGGACGGGCTCCTTCAATTAGCACGGGGAGCTTCCTG 9848
Qy      471 ------------------------------------------------------------ 471
Db     9847 GCCTATTATGGAATCAATGCAATAAGTAGACTAGATGGCAGAGGCATTTCTATCCTTCAC 9788
Qy      471 ------------------------------------------------------------ 471
Db     9787 CCTAGATTGTCTGTAACAATTTAGCTATACATAGACCAGTAGGATCCCCAGACAGTCTCC 9728
Qy      471 ------------------------------------------------------------ 471
Db     9727 AGCTAGTCAGACATGGAGTGGGCTGTGACAGTCCCTTTGTCCCCCCCGTGCTACCACCTC 9668
Qy      472 -------AspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnVa 489
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9667 AGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGT 9608
Qy      489 lArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGl 509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9607 TCGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGG 9548
Qy      509 yAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAsp-------------- 524
            ||||||||||||||||||||||||||||||||||||||||||||||
Db     9547 GGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGGTAGGACCTCTGA 9488
Qy      524 ------------------------------------------------------------ 524
Db     9487 ACCTGCAGTGACCTGGGACCCAGAGACACTCCCACTCAGGGACAGTGTGGCCCAAGGCCC 9428
Qy      524 ------------------------------------------------------------ 524
Db     9427 TCTTCCCTGCCTTGTGATGTCTGTTGTGCCCTATAAAGCAGTGGGCAGAGGTTCTTCTTC 9368
Qy      524 ------------------------------------------------------------ 524
Db     9367 CCAGTGTAAAATGGGGAAACTGGGGCTCAGAAAATGAAAATGACTTGCCAAGCTCAGCAG 9308
Qy      524 ------------------------------------------------------------ 524
Db     9307 GGCAGCTGGGACTTGAGCCCAGCTCTGAGCCCCCATACCATGCAGCCCTTCTTCCTAACA 9248
Qy      524 ------------------------------------------------------------ 524
Db     9247 TACATTTACTACCCAGCCATTGGATATGAAAGCTTCTGGCCCACACTGTCCACTAGCCTT 9188
```

```
Qy        524 ------------------------------------------------------------ 524
Db       9187 GGGATAGAAAGGATGTCTCCGCATCCTCCAGCGTAGGGCTCCCCTGAGAGGCTCCCTCCC 9128
Qy        524 ------------------------------------------------------------ 524
Db       9127 TCCCTCCCTCAGATCCTGGGCCTGGGCCTACTTCTGACTTGCTGTGTGACCTGGGCAAGT 9068
Qy        524 ------------------------------------------------------------ 524
Db       9067 CTCTGATCACTCCTGACTGTGTCTTATCTTCCCTGTTGGTGGAGCGGAGCTCACGCCAAT 9008
Qy        525 ------------ValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProS 540
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       9007 GTTCTTCTGTCTCAGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCT 8948
Qy        540 erAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuL 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8947 CGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGA 8888
Qy        560 ysTyrLysIleGluTyrGlyLeuGlyLys------------------------------- 569
              |||||||||||||||||||||||||||||
Db       8887 AGTACAAGATAGAGTACGGTTTGGGGAAGGAAGGTGAGTGGGGCCGGGGAGGTACAGAGA 8828
Qy        569 ------------------------------------------------------------ 569
Db       8827 TGTCAGTGGGAATGGGGGTCCTCCCCTAGATGAGGGACGTTGTCAGTGGAAGTGGAGCCT 8768
Qy        569 ------------------------------------------------------------ 569
Db       8767 AATTGGGGCATGGGATGAGAAAGCTCTGGAAAAGAGCATTGGGAGACTAGATAGCCCAAT 8708
Qy        569 ------------------------------------------------------------ 569
Db       8707 GTGTAAAGTATGTGTGCAAGCCCAAAGGCCTGAGTTCAATCTCAGACTTCATGTTAAAAA 8648
Qy        569 ------------------------------------------------------------ 569
Db       8647 AAAAAAAGCCTGGCACAGTGGTAGGTGCCTGCTTGTAATCCCAGAGCAGGGGAGGTGGAG 8588
Qy        569 ------------------------------------------------------------ 569
Db       8587 ACACGGGGACACCTTGAGCTCACTGGGCTAAACCAGTGAGTTCCAGGCCAGCGAGAGTCC 8528
Qy        569 ------------------------------------------------------------ 569
Db       8527 CTGTGCAGAAGAGAAGGTGGACAATACCAGAGAAATGACAACAGAGGTTGTCATCTGACC 8468
Qy        569 ------------------------------------------------------------ 569
Db       8467 TCCACATGTATGTGTACACACACCTGCACATAGGTGAACACTCACACTCATGGACACACA 8408
Qy        569 ------------------------------------------------------------ 569
Db       8407 CGGACACAGGGTTTCGGGGCAGTTGAGGAAAAGGGAAAGCAAAGAGGGTCAGTAGCCTGC 8348
Qy        569 ------------------------------------------------------------ 569
Db       8347 GCACAGATAAGGAGGGTGATCAGTGGGACCTTGGACTGTTCCCTTTCTGTGAAGGGAGCA 8288
Qy        569 ------------------------------------------------------------ 569
Db       8287 ATTTACCCAGTGGATCCTTGGAACCCTTCTAGGTTTTTGTGTTTGTTTATTTTTTAAGAT 8228
Qy        569 ------------------------------------------------------------ 569
Db       8227 GTATTTATTTATTTTATGTATATGAGTTCTCTATCTGCACGCCCACCTGCATGCCAGAAG 8168
Qy        569 ------------------------------------------------------------ 569
Db       8167 AGGGAATCAGATCCCAGTACAGATGGTTGTGAGCCACAGTGTGGTTGCTGGGAATTGAAC 8108
Qy        569 ------------------------------------------------------------ 569

Db       8107 TCAGGACCTCTGGAAGAGCAGTCAGTGCTCCTAACAATTGAGCCATCTCTCCAGCCCCAG 8048
Qy        569 ------------------------------------------------------------ 569
Db       8047 AAACCTTCTGGTTTACACACCAGAGAAGATGTGGGAGGCTGCAGGAGTGCACATGAGACA 7988
Qy        569 ------------------------------------------------------------ 569
Db       7987 TGGGCTGCTCCTACTGGGGAGATGATGCTGGTCTGGGGGAGGTGATGTGGTTGGGAAAGG 7928
Qy        569 ------------------------------------------------------------ 569
Db       7927 AATGGAGAGTGAGGAAAAGAAAAGTGGTGAGGGTCAGGGTAAAGCTAAAGAGAGAGGCAG 7868
Qy        569 ------------------------------------------------------------ 569
Db       7867 GAGGCTGATGGGAGACAGCAGTCAGGGCTCCTGACACCAGGGACTGCTTAGTGTCCTTAG 7808
Qy        569 ------------------------------------------------------------ 569
Db       7807 CCAGCCATCAAGCCTGTGTAAGAGGTGTGTACTGGCTGTTGCTGTTGGTTCACAGGGAGG 7748
Qy        569 ------------------------------------------------------------ 569
Db       7747 AAAAGTTCATGCCTTGGTTTCTCCATATGGAACGGAGGCCAATGGGCTTCAGCAGCGATA 7688
Qy        569 ------------------------------------------------------------ 569
Db       7687 GGTAGAAGCCCAGGGTCAACCTGGAGGCATGAAACCTGCCTCCGTGCTGAGGGGTCAGTG 7628
Qy        569 ------------------------------------------------------------ 569
Db       7627 CCCAGAGGACCCTGGGGACCTGAAATGGAAAAGTCAATGTGATAAATTTAAAGATTGATG 7568
Qy        570 --------------------------------------------GluAspGlnValPheS 575
                                                          ||||||||||||||||
Db       7567 GTCCCCCAATGAAGGACACCCAGCAGCCTATCCTTTTTGTCCCAGAAGATCAGGTTTTCT 7508
Qy        575 erThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysV 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7507 CCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAG 7448
Qy        595 alTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpM 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7447 TGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGA 7388
Qy        615 etGlnHisArgThrProGlyValHisAsnGlnSerHis---------------------- 627
              ||||||||||||||||||||||||||||||||||||||
Db       7387 TGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGGTATGGACCTACACGGGCAGA 7328
Qy        627 ------------------------------------------------------------ 627
Db       7327 CTGGACAGCGGGGACAGGTGCTCTGTGGCCTTAGATATTGCCCCTGTTATTGCCTGTTAT 7268
Qy        627 ------------------------------------------------------------ 627
Db       7267 TGCCACCCCCCCCCCCCACAAACACACACTTGATGCACCCGGAAGGCCTGGGTGTAGGA 7208
Qy        627 ------------------------------------------------------------ 627
Db       7207 GTGAAGGGCTGGGCTCAAATATCTCCTTGTTCACTCTGTGACTTTGACAAGCCACTCGCA 7148
Qy        627 ------------------------------------------------------------ 627
Db       7147 TCTCCAGAGCTTCGGTACTACCAAGGATGATAGCAGGACTAGTCATTGTAATTATACACA 7088
Qy        627 ------------------------------------------------------------ 627
Db       7087 GAGTGGTCGGCATAGAACAGGAGCTCCTTGGGTTGAGACCACGGCTCACTTGACAGAGTG 7028
Qy        627 ------------------------------------------------------------ 627
```

```
Db   7027 CTTGCCTAGCATGTACAAAGCACTAAATTCAATCCCCAGGACCACATAAACCCAGCATAC 6968
Qy    627 ------------------------------------------------------------ 627
Db   6967 CTCTGATCCTGGCAGTTCAAGGTCTGCCTCACCTACCTAGTGAGTTGGGGACCAGCCTAG 6908
Qy    627 ------------------------------------------------------------ 627
Db   6907 GAGGTATGAAACTCTGTATCAAAGAAAGGAAGGAACCCATTAAAAGCTAGCTGGCTTTGC 6848
Qy    627 ------------------------------------------------------------ 627
Db   6847 CAAGAGGCCCAAACAGTAGAGTGGTTAAATGGCTACAAAGTCTCCAGGCAGGTGCTAGTG 6788
Qy    627 ------------------------------------------------------------ 627
Db   6787 GAGGCTCATGCCTTTAATCCCAGCACTCGAGAGGTGGAGGCAGGCAGATTTCTGAATTTG 6728
Qy    627 ------------------------------------------------------------ 627
Db   6727 AAGACAGCCTGGTCTACACCAAGGTTACACAGAGAAACCCTGTCTCAAAAAACCAGAAAA 6668
Qy    627 ------------------------------------------------------------ 627
Db   6667 GAAAAAAGACTACAAAGTCAGCTACCATATTCAGACATTAGACAAGCTATAGCTTAAAAA 6608
Qy    627 ------------------------------------------------------------ 627
Db   6607 AAAAAAAAAAAAAAAGCTGGATTATGGAATTATAAGAATGGGCCGGGCGCGGTGGCGCAC 6548
Qy    627 ------------------------------------------------------------ 627
Db   6547 GCATGCCTTTGATCCCAGTACTCGGGAGGCAGAGGCAGGCGGATTTCTGAGTTCGAGCCT 6488
Qy    627 ------------------------------------------------------------ 627
Db   6487 GGTCTACAAAGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTCGAAAAA 6428
Qy    627 ------------------------------------------------------------ 627
Db   6427 AGAAAAAGAAAAAGAAAAAAAAATGAATGAAGTAAAAATATACAAAGGCTGCCAAGGTCG 6368
Qy    627 ------------------------------------------------------------ 627
Db   6367 CCCAGCTAATAAAGGGGCTTGCTCCATTTTCTGGGGATCTGACTTCCAATGTCTCCCCAG 6308
Qy    627 ------------------------------------------------------------ 627
Db   6307 AATTTATATGATGAGTGAGTCTACAGCTTGCGGCCTCTGACCGCCACCAGCACGCACATA 6248
Qy    627 ------------------------------------------------------------ 627
Db   6247 AATTAAATAAATACTTTAAATTTTTTAAAATACATATACAAAATGTTTGAAATGGGGTCT 6188
Qy    627 ------------------------------------------------------------ 627
Db   6187 GACACACAGCCAACCCTCTACTCTTGGGCTCCGAGTGTTGCTCGTCGTGTTGATGGTCCC 6128
Qy    627 ------------------------------------------------------------ 627
Db   6127 AAATCTCCCTCATGATCTCAGCAGCCCCTTGACCATACTAACTCTTTCAAGAGACCTGCA 6068
Qy    627 ------------------------------------------------------------ 627
Db   6067 GAAGCCCCTCTGCCCCTGACCGCGGGCCTGCGCAGACACGCCCCCTGCTGTCTCTTGTGC 6008
Qy    627 ------------------------------------------------------------ 627
Db   6007 GTGCCCGGCAGACGCGCCTGCGCAGCTAACATGCAGAGCTCAAACCCACCGGGAAGTGGC 5948
Qy    627 ------------------------------------------------------------ 627
Db   5947 AGAGTGGAGCAGTTAGGGAAAGCCCAGGCACAGCTCTATGCTGTGCTCCGCCCCTGCCCT 5888

Qy    627 ------------------------------------------------------------ 627
Db   5887 GCACCGGCTCACCTCCCTGACCACTTCACTCCTCATTCTCTTCCCCTTGTCCCCATCATC 5828
Qy    628 ---------------ValProPheAlaProAlaGluLeuLysValArgAlaLysMetGlu 642
                         |||||||||||||||||||||||||||||||||||||||||||||
Db   5827 TTCCTCTGCCCCCCAGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAG 5768
Qy    643 SerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeu 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5767 TCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTC 5708
Qy    663 TyrTrpGlyGluValGlyThrGluGluAlaAspGlyAspArgProProGlyGlyArg 682
          ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5707 TACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGT 5648
Qy    683 GlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGlu 702
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5647 GGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAA 5588
Qy    703 LeuThrGlnLeu------------------------------------------------ 706
          ||||||||||||
Db   5587 CTGACCCAGTTAGGTGAGTAGGCTTGGGCTTGGGAGGGTCATGGGCAAGCTAAGCACACG 5528
Qy    707 ---------------------------ValProGlyArgProTyrGluValLysLeuVal 717
                                     ||||||||||||   ||||||||||||||||||
Db   5527 GAGAGTGACACTCATGGCCTTTCCCCAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTA 5468
Qy    718 AlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAla 737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5467 GCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCG 5408
Qy    738 ProThrProAsp------------------------------------------------ 741
          |||||||||
Db   5407 CCCACGCCAGG-TGAGGGGGAGGCAGGCGGGGGTCTTGGAGCCTTCTTTTCCTTTTTCTT 5349
Qy    741 ------------------------------------------------------------ 741
Db   5348 TATCAGCCAGCAAGCCGTGTGCATGCCACAGCCACCCGGCCCACCCTCCTGGGCAGAAAT 5289
Qy    741 ------------------------------------------------------------ 741
Db   5288 CCTCCTGCTCTAGACATGCCCCTCATTCTCAAGCTAAGCAGAGCGAACGGGAGAGATTGG 5229
Qy    741 ------------------------------------------------------------ 741
Db   5228 AAAGAGAAACTCCGTGTCCGATGCGTGATTCCTGGCTCAAGCTTTAACCTTTAAACAGCT 5169
Qy    741 ------------------------------------------------------------ 741
Db   5168 TTTTAGATCGCATCCACGTATACCACATGGTCCGTCAGATTATCCCATAATTCTTTTTTT 5109
Qy    741 ------------------------------------------------------------ 741
Db   5108 TAAAGACTTACTTACTTATTTTATGTATATGAGTACACTGTAGCTGTCTTCAGACACACC 5049
Qy    741 ------------------------------------------------------------ 741
Db   5048 AGAAGAGGGCACCAGATCCCATTACAGATGGTTGTGGCTTCTGGGAATTGAACTCCAGAC 4989
Qy    741 ------------------------------------------------------------ 741
Db   4988 CTCTAGAAAAGCAGTCAGTGCTCTCAACCACTGAGCCATCTCTCCAGCCCCACCCCATAA 4929
Qy    741 ------------------------------------------------------------ 741
Db   4928 TTCTTAAGCAGGGGCAAGCACAACACAGTGGTTTGTCACCCTGACGATCCAGCCTGACAG 4869
Qy    741 ------------------------------------------------------------ 741
Db   4868 TCAGAAGAAACCTGGGTTCTAATTCCAGCGAGTTCAGTTTTTCATTGAGCAAATATTTT 4809
```

```
Qy       741 ------------------------------------------------------------ 741
Db      4808 ACCTAAGTATCTACCACATGACAGCTCTGTTCTGGAAACTTAGTAATGAACAACATGAAT 4749
Qy       741 ------------------------------------------------------------ 741
Db      4748 AAGGTATCTGCTCTCTTGGAGCTTATTTTATGCTTATCTAATAAACATTTGTCATCACCA 4689
Qy       741 ------------------------------------------------------------ 741
Db      4688 AATAGTATACACTGTCCCACAGTCTTCTACCTGCTTTGTAAGCAGCCTCCTAAGGACACT 4629
Qy       741 ------------------------------------------------------------ 741
Db      4628 CATGAGGGGACAGTGAGATGGTTCAGCAGGGAAGGTGATTGCTACAAGCCACATGACCTG 4569
Qy       741 ------------------------------------------------------------ 741
Db      4568 AGCTTGGTTTCTGCACCCCCCCCCCAAATTGTCCTCTGCTCCCCCACATGCACACAGCAG 4509
Qy       741 ------------------------------------------------------------ 741
Db      4508 CACACTCCCTCCCTTCCCTCCACACAATAAATGTGTGACTGAGAAGATGCAGGAGTTAAG 4449
Qy       741 ------------------------------------------------------------ 741
Db      4448 AATTTTAAATATGTTTTAAAGATGTATTTATGTATTTATATGTGTGAGTACACTGTCACT 4389
Qy       741 ------------------------------------------------------------ 741
Db      4388 GTTGCTATCTTCAGACACAGCAGAAGAGGTCACTAGATCCCATTATAGATGGTTGTGAGC 4329
Qy       741 ------------------------------------------------------------ 741
Db      4328 CACCATGTGGTTGCTGGGAATTGAACTCAGGGACTCTGGAAGAGCAGTCAGTGCTTAACC 4269
Qy       741 ------------------------------------------------------------ 741
Db      4268 GCTGAGCCCTCTCTTCAGCCCCCAGAATTTTAAATATTTTAAAGAAGTAAATATAAAGAA 4209
Qy       741 ------------------------------------------------------------ 741
Db      4208 AAGAAGAAAAAATGTCAAGATGTAGGTATTGTCATCAGATCCTTATAAGAATAGAAAATGA 4149
Qy       741 ------------------------------------------------------------ 741
Db      4148 TGGCACAGAGAGGTTAAACAACTTCTCTAAAGCCACATACATAGCTAGGAAGTGGCAGCA 4089
Qy       741 ------------------------------------------------------------ 741
Db      4088 ATGGAGCCTGGCTCCAGAGTCAGGGCATGCTTTCTCTGTCACATTCTCATCAGCCATGGG 4029
Qy       741 ------------------------------------------------------------ 741
Db      4028 ATATGTCACAGTCACTCTATGTGTGACCTAAGCAAAAAAAAATCACTTCCATTCTTGGGA 3969
Qy       741 ------------------------------------------------------------ 741
Db      3968 GCCTCAGCTGCAGTAGGAATAATGTCACCCAACTCACAGGATCACAAGAGAGGCCCGGTG 3909
Qy       741 ------------------------------------------------------------ 741
Db      3908 CCAGCTGTGGTATCACACAGCTGTAACCCCTGTGCTAAAGGACCAGAGGCACAAGGATGG 3849
Qy       741 ------------------------------------------------------------ 741
Db      3848 AGAATTACAGGCCAGCCTAGGCTACTACATAATGAGATTCTGCCTCAAAAATAAGAGGGG 3789
Qy       741 ------------------------------------------------------------ 741
Db      3788 GAAAGGCCAAATGAAACCTACTGAGACCATGTACGTCCACCTACTCTCTGCACACTTGGG 3729
Qy       742 ---------------LeuProIleGlnArgGlyProProLeuProProAlaHisValHis 756
                            |||||||||||||||||||||||||||||||||||||||||||||
Db      3728 TTTCCCCCTCTAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCAC 3669
Qy       757 AlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThr 776
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3668 GCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACT 3609
Qy       777 ValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeu 796
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3608 GTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTG 3549
Qy       797 ValThrTyrTyrThr--------------------------------------------- 801
             |||||||||||||||
Db      3548 GTCACCTACTATACCAGGTATCAGTGAGGGGGGAGTGTAGCTGGGGTGGTAGACTAGCGG 3489
Qy       801 ------------------------------------------------------------ 801
Db      3488 GTTCAGCTGTGGTGAGGCCACACCCAGGATCTACACAGATGTCTTTTCCGTGACTGCTGA 3429
Qy       801 ------------------------------------------------------------ 801
Db      3428 CCTGGGGGGTAATAATGAGGGATGGGATTCCCCCATCCCTCAGTGGCAGAATGTGTGGTC 3369
Qy       801 ------------------------------------------------------------ 801
Db      3368 AGCCTTTGGTGAGAGTGGATGCAAGACTAGAGGCTTCCATCCGTTGGCTGGCACCCTGCC 3309
Qy       801 ------------------------------------------------------------ 801
Db      3308 GGCCCATCATTCGGAAGATAGCTTAAGGAGCTGCTGTTCTGATTGGGAAGACACCCTGTG 3249
Qy       801 ------------------------------------------------------------ 801
Db      3248 AAAGCCAACAGAGAGAGAAACTAAGCTGAGGATAGACACTTCAGGAGCGATTGGTCCCCC 3189
Qy       801 ------------------------------------------------------------ 801
Db      3188 AGGTGCAGAATTACTTAGATTAATTCGGGAAAGCGTCCCAGAGGAGAGAAGGGCTTGCTT 3129
Qy       801 ------------------------------------------------------------ 801
Db      3128 TGGAGATGAGACATGGAATAACATATGTCCGGCCTTTCTGTGGTTTGCCTGGCTGTGTGT 3069
Qy       801 ------------------------------------------------------------ 801
Db      3068 GGGAGGGGAGGTTGTGGGCCCAACAAGCAGCTCCCTCCAGGGTGGGGCCTGGCTGGAGGC 3009
Qy       801 ------------------------------------------------------------ 801
Db      3008 AGGGGCTTCTGTCTGGGGTTCTGCCTGGCTTGAGGCCTGTCTGAGGGCACTAGTCCTATC 2949
Qy       802 ------------SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLys 817
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db      2948 CCTTCTCCTGGAAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAG 2889
Qy       818 TyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValVal 837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2888 TACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTA 2829
Qy       838 GluArgSerThrLeuProAspArg------------------------------------ 845
             ||||||||||||||||||||||||
Db      2828 GAACGCTCCACCCTGCCTGACCG-TGAGTATCCACTGCCCGCTCCCCACCTGTGCCACCA 2770
Qy       845 ------------------------------------------------------------ 845
Db      2769 TTTGCTGGGTCTGAGGAAGGAAGAAGAAGAGGAGGCTGTGCCCTCATCCCCTTAGTTAGG 2710
Qy       846 ------------------------------------------ProSerThrProProSer 851
                                                       ||||||||||||||||||
Db      2709 GCTGTGTCTCTGACTCTCCACCCCTTCCCCTGTGCTGCAGGGCCTTCAACACCTCCTTCT 2650
Qy       852 AspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThr 871
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2649 GACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACG 2590

Qy        872 GluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnPro 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2589 GAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCC 2530

Qy        892 GluHisGlnTrpThrLeuLeuThrThrGlu------------------------------ 901
              ||||||||||||||||||||||||||||||
Db       2529 GAACACCAGTGGACACTGCTCACCACAGA-GGGTAAGGACGCCCACCCCTGCCTCCTTGC 2471

Qy        901 ------------------------------------------------------------ 901

Db       2470 CCTTTTGTTCCAAGGGTCTCTGGGATTGGAGGCTGAAGCTACTGTCTGGGTGCAAAGTTG 2411

Qy        902 ---------------------------------------------------------Gly 902
                                                                       |||
Db       2410 GTTTGCCCATGGTGCTTTGGCCTCCTGCCACCCTGTCTCCTCCCACTTCCCACCACAGGA 2351

Qy        903 AsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMet 922
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2350 AACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATG 2291

Qy        923 GlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeu 942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2290 GGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTG 2231

Qy        943 GlnGluThrPheSer--------------------------------------------- 947
              |||:::|||||||||
Db       2230 CAAAAGACATTCTCAGGTACCAGGCAGGAGGGAGAAGGCCCTTGGGACAACTGGAGGCCC 2171

Qy        947 ------------------------------------------------------------ 947

Db       2170 CAAGACCCCTGGATTGGCTCCCAAGGCCCTGCCTTCTGGAAATTACATTTTCTTTACTTT 2111

Qy        947 ------------------------------------------------------------ 947

Db       2110 GGAACCCTAAGGTGTGGGGGATGGAGAAGCAGTGACAGCTTGATAGAGGGACTGTCATAG 2051

Qy        947 ------------------------------------------------------------ 947

Db       2050 CTTTGAGGGGTCTTTTATTATTTTATTCTCTCTGTGTGTGTTTGGCACAGTGCATATGCG 1991

Qy        947 ------------------------------------------------------------ 947

Db       1990 AAGGTCAGAGGACATCAGAGAGTTGGTTTTGTGGGGATCTTGCTTTGCCTGGAGGGTTGG 1931

Qy        947 ------------------------------------------------------------ 947

Db       1930 GTGAGCTCCTGAGTTCTGTGAAGAAGTGGTCAATAGGCTGGATAGACACTCACAAGGGCT 1871

Qy        947 ------------------------------------------------------------ 947

Db       1870 CAATTAAGATACAAAATGAGATCCCAGGGCTCCTTATTGTAAATTATTGAAGATGTGAAG 1811

Qy        947 ------------------------------------------------------------ 947

Db       1810 ATGAGCCAGTGGGGCATTAGCCACACATAGGACTGAGGAGCATGAAGCCCTGAGTCCAAA 1751

Qy        947 ------------------------------------------------------------ 947

Db       1750 TGCACACACACACACACACACACACACACACACACACACACCGTAAAACTTGGCTGTGGC 1691

Qy        947 ------------------------------------------------------------ 947

Db       1690 TGCTGTAATCCTTGCTGAGGATGCAGAGACAGGATATTTCTGGGGCTGAGAGGGACTGGG 1631

Qy        947 ------------------------------------------------------------ 947

Db       1630 GAAGTCACGTGATGTAGACCTCTGATCTCCACACTACATATACATGCCCTTAGATGAACA 1571

Qy        947 ------------------------------------------------------------ 947

Db       1570 TATACAGACATGACACACACATGTGTGCACCTAGAGCCCACCTTGGTTATACATATATAC 1511
```

```
Qy        947 ------------------------------------------------------------ 947

Db       1510 ACCATGAGAGCCCTTGTTGCCAGGAGGAAAACAACTTCACTGAAGAGCTGATTCTATAGA 1451

Qy        947 ------------------------------------------------------------ 947

Db       1450 GTCCAGTTTTGTGAGGAGTGGCAGGTTCTGTACAGCAGGGAAGACCGTGGTGGACTCAGT 1391

Qy        947 ------------------------------------------------------------ 947

Db       1390 ATGAGGCCCAGTAAGATCCCGTGGCTCTCTCTATTCCAATCAGAAGGGAGTGAGCTTCCA 1331

Qy        947 ------------------------------------------------------------ 947

Db       1330 AACCAGGTGTGAACCAGGTGTGGTACTGCACACCTTTGATCTCCGTCCTGGTTTATGAGC 1271

Qy        948 ------------------------------------------AspSerLeuAspValHis 953
                                                        ||||||||||||||||||
Db       1270 ACTGCTTCCTTCTGACCTGGAGTCACACTGTTCTCCCCCGCAGACTCCTTGGATGTGCAC 1211

Qy        954 AlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMet 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1210 GCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATG 1151

Qy        974 CysAlaGlyLeuArgGlnSerSer------------------------------------ 981
              |||||||||||||||:::||||||
Db       1150 TGTGCTGGCCTACGACGAAGCTCCCACAGGTGAACCGGGGAACCACTGAGACCACTGAGT 1091

Qy        981 ------------------------------------------------------------ 981

Db       1090 TACCAGGAAGGGCTGCAGAGTGGACAAAGGAGCGTGTCCTTCCCTGCTGCCCCGGAAGCC 1031

Qy        981 ------------------------------------------------------------ 981

Db       1030 CCAGTATAGATAGGGCTATGGAGCTGGAGTGAAGTTCAGTGCTACATGTGCCCCCATGTG 971

Qy        981 ------------------------------------------------------------ 981

Db        970 AGGCCCTGGGCGTCATCCTCAGTAGCACAAAACAAGTGGATAATTTATTTATTAGGATTC 911

Qy        981 ------------------------------------------------------------ 981

Db        910 TGAGATGAAGGATTGAGGTGTTGTTCTCATGAGAATGTAGAATAAACTTAGGGAAGCCCC 851

Qy        981 ------------------------------------------------------------ 981

Db        850 AGGACTCCCCACCTCAGTAACAGCGGGCTGAGCTACCTGTGGGCTTCCCTCATGGCTTAA 791

Qy        981 ------------------------------------------------------------ 981

Db        790 ATGGAAATGGCCCATCTCGCCATACAAGGAAGCGTGCCATCTGAGGGGTCAGGCCGAGTG 731

Qy        981 ------------------------------------------------------------ 981

Db        730 GACTTGTATTCCTTCCTTTCTGTTTGATTGCACCACACCAGGAAACCAAATCCACTCTTG 671

Qy        982 ---------------------------HisArgGluAlaLeuProGlyLeuSerSerSer 992
                                         |||||||||||||||||||||||||||||||||
Db        670 TCCGACCCATAACTTTGTTCTGTTGATCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCA 611

Qy        993 GlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValPro 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCT 551

Qy       1013 AlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAspTrpSerProProPro 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 GCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCC 491

Qy       1033 SerAspValGluAspLysAlaGluValHisSerLeuMetGlyGlySerValSerAspCys 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 TCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGC 431
```

```
Qy       1053 ArgGlyHisSerLysArgLysIle-SerTrpAla------GlnAlaGlyGlyProAsnTr 1070
              ||||||||||||||||||||:::     |||:::       ::::::|||
Db        430 CGGGGCCACTCCAAGAGAAAGGTGAGACTGGAGTCTCTTCAGGTCAGGG----------- 382

Qy       1070 pAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuTh 1090
                                               :::   |||   |||        ||
Db        381 ------------------------------ACTCACACGAACCCATTCCCAAATCAAAC 353

Qy       1090 rArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTy 1110
              |               ||||  ||||||                 :::|||   :::
Db        352 A---------------CCAGGGGGCACC-----------------AGAGCACAACTTCT 326

Qy       1110 rAspGly---IleLysSerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnVa 1129
               ||||||    ::::::|||
Db        325 -GACGGTGAGATGCGCTCGCGCGCACACACACACACACACACACACACACACACACACAC 267

Qy       1129 lGluAlaGluValIleValHisSerAsp------------------PheGlyAlaSerLy 1143
                     :::   :::    |||::::::                  |||   |||::
Db        266 ACACACACAGGCACTGAACCATACAGAGAAAGCCTGTCCATACCACCTGGGAGGATCCCA 207

Qy       1143 sGlyCys-----------------ProAspLeuHisLeuGlnAspLeuGlu 1154
              :    |||                       |||||||||   ::::::|||
Db        206 GAGCTGTGTTTCGAGAGCACCCATACTTCTCTTCACTTACCTAATATGGAG 155
```

Search completed: October 14, 2003, 08:42:16
Job time : 8508 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         October 14, 2003, 06:11:05 ; Search time 3845 Seconds
                                             (without alignments)
                                             7913.967 Million cell updates/sec

Title:          US-09-754-997A-2
Perfect score:  6604
Sequence:       1 MARADTGRGLLVLTFCLLSA..........CPLTVSPSLPRAPVSSAQVP 1252

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09754997/runat_14102003_061100_9718/app_query.fasta_1.1415
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09754997_@CGN_1_1_2865_@runat_14102003_061100_9718 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2595	39.3	2626	11	AK079073	AK079073 Mus muscu
	2	1610.5	24.4	4185	11	AK083540	AK083540 Mus muscu
	3	1326	20.1	900	12	BI185886	BI185886 UNL-P-FN-
	4	1292	19.6	817	14	CA317106	CA317106 UI-M-FW0-
	5	1252	19.0	740	13	BU701299	BU701299 UI-M-EX0-
	6	1216	18.4	699	14	CD351162	CD351162 UI-M-FY0-
	7	1204	18.2	713	14	CB524810	CB524810 UI-M-FY0-
	8	1202	18.2	741	14	CD352400	CD352400 UI-M-GI0-
	9	1174	17.8	719	14	CA751384	CA751384 UI-M-FO0-
	10	1098	16.6	631	14	CB518504	CB518504 UI-M-GH0-
	11	1097	16.6	759	14	CB988160	CB988160 AGENCOURT
	12	1082.5	16.4	703	14	BY712424	BY712424 BY712424
	13	1047	15.9	591	13	BU057120	BU057120 UI-M-FO0-
	14	1046	15.8	2906	11	AK051027	AK051027 Mus muscu
	15	1032	15.6	651	10	BB624536	BB624536 BB624536
	16	1008	15.3	637	10	BB665894	BB665894 BB665894
	17	1002	15.2	630	10	BB626002	BB626002 BB626002
	18	1000.5	15.1	3230	11	AK011916	AK011916 Mus muscu
c	19	990.5	15.0	717	13	BX100607	BX100607 BX100607
	20	977	14.8	553	14	CA316452	CA316452 UI-M-FW0-
	21	971	14.7	632	10	BB649997	BB649997 BB649997
	22	960	14.5	544	14	CA892088	CA892088 B0170A06-
	23	945.5	14.3	1137	11	AK082162	AK082162 Mus muscu
	24	927	14.0	578	14	CB607425	CB607425 AMGNNUC:N
	25	910	13.8	595	12	BM256356	BM256356 518704 MA
	26	882	13.4	504	14	CA876082	CA876082 K0945H09-
	27	879	13.3	509	14	CA873325	CA873325 K0924D02-
	28	803	12.2	455	14	CA883713	CA883713 B0105A08-
	29	784	11.9	442	14	CA891757	CA891757 B0167G06-
	30	778	11.8	435	14	CA892100	CA892100 B0170B03-
	31	753	11.4	672	9	AW143015	AW143015 EST293310
	32	725	11.0	747	13	BU382531	BU382531 603858107
	33	708	10.7	438	14	CB751460	CB751460 AMGNNUC:N
	34	700.5	10.6	836	14	CB994721	CB994721 AGENCOURT
	35	689	10.4	524	9	AA947283	AA947283 ok20b04.s
	36	687	10.4	564	28	AZ354558	AZ354558 1M0093E22
	37	682	10.3	466	9	AI368698	AI368698 ta37b09.x
	38	678	10.3	400	14	CB699715	CB699715 AMGNNUC:S
	39	671.5	10.2	411	14	CA871247	CA871247 K0908H04-
	40	667.5	10.1	3642	11	AK084609	AK084609 Mus muscu
	41	638.5	9.7	485	9	AI741225	AI741225 wg26e07.x
	42	601	9.1	353	14	CA886377	CA886377 B0126D12-
	43	588	8.9	624	13	BU056331	BU056331 UI-M-FO0-
	44	575.5	8.7	500	12	BI182798	BI182798 UNL-P-FN-
	45	564	8.5	939	13	BQ646105	BQ646105 AGENCOURT

ALIGNMENTS

```
RESULT 1
AK079073
LOCUS       AK079073                2626 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male diencephalon cDNA, RIKEN full-length
            enriched library, clone:9330155G14 product:hypothetical protein,
            full insert sequence.
ACCESSION   AK079073
VERSION     AK079073.1  GI:26098261
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2626)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2626
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9330155G14"
                     /db_xref="taxon:10090"
                     /clone="9330155G14"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1..2626
                     /note="hypothetical protein (evidence: rsCDS)"
BASE COUNT      599 a    767 c    724 g    536 t
ORIGIN

Alignment Scores:
Pred. No.:              5.21e-165       Length:         2626
Score:                  2595.00         Matches:        563
Percent Similarity:     84.66%          Conservative:   22
Best Local Similarity:  81.48%          Mismatches:     60
Query Match:            39.29%          Indels:         46
DB:                     11              Gaps:           11

US-09-754-997A-2 (1-1252) x AK079073 (1-2626)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         15 ATGGCGCGGGCGGACACGGGCCGAGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 74

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 134

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 194

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 254

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 314

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 374

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 434

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 494

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
```

```
Db      495 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 554

Qy      181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
Db      555 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 614

Qy      201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
Db      615 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGCCCACCAATTCAGCCCCCCAACGA 674

Qy      221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
Db      675 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTTAAAAGGTCTTTGGGGGGTAACAAGGGA 734

Qy      241 GlnAspValValIleVal-AlaAlaProGluAsnThrThrValValSerGlyGlnAsnVa 260
Db      735 CAGGATGTGGTCATTGTGCCCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGT 794

Qy      260 lValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAs 280
Db      795 AGTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGA 854

Qy      280 pGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSe 300
Db      855 TGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCCCCAG 914

Qy      300 rAlaGlnProArgHisSerGlyVal-TyrValCysArgAlaAsnLysProLeuThrArgA 320
Db      915 CGCCCAGCCTCGGCACTTTTGAATTTAATTTTTCCCAACCCACCAACCCCCCCACGCCTT 974

Qy      320 spPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle-SerGlnAla 339
Db      975 AATTTCCCCCTTCCGCTTCTTAAGTTCCAATTCTTTCTTCCCCCACCCTTTTCCCAGCCC 1034

Qy      340 ProGluAlaLeu-SerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGl 359
Db     1035 GCCCAAGCCCTTTTCCCGAACCCGGCCCACCCCCCCCCGTTTCTTTTCCCGGCCTTCCGG 1094

Qy      359 yGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyAr 379
Db     1095 GGACCCCCGGCCCGCGCTGCACTGGCTGCACGACGGAATCCCGTTGCGACCCAATGGGCG 1154

Qy      379 gValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAl 399
Db     1155 CGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGC 1214

Qy      399 aGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLe 419
Db     1215 TGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCT 1274

Qy      419 uAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLe 439
Db     1275 GGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCT 1334

Qy      439 uSerSerSerSerValLeuValAlaTrpGluArgProGlu-LeuHisSerGluGlnIleI 459
Db     1335 GAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTTGCACAGCGAGCAAATCA 1394

Qy      459 leGly-PheSerLeu-HisTyrGlnLysAlaArgGlyValAspAsn-Val-GluTyrGln 477
Db     1395 TTGGCTTTCTCTCTTTCACTACCAAAAGGCAAGGGGAGTGGACAATGGTGGGAGTACCAG 1454

Qy      478 PheAlaVal---AsnAsnAspThrThrGluLeuGlnValArgAspLeuGlu--------- 493
Db     1455 TTTGCAGTAAAACAAAATGGGAAAACCAAA---------AGAAGCTTGGAAGGGTTTCGG 1505

Qy      494 ---------------ProAsnThrAsp---TyrGluPheTyrValValAlaTyr---Ser 506
Db     1506 GGAACCTGGGAAAACCCAACCCCGGATTAATGAAGTTTTAAGTGGTTGGCCTAACTTCC 1565

Qy      507 GlnLeuGlyAla---SerArg-ThrSerSerPro---AlaLeuValHisThrLeuAspAs 524
Db     1566 AAGCTGGGGGGGCCAGGCCGAAACCTTCAAGCCCAGGCCCCTGGTGCATACACTGGACGA 1625

Qy      524 pValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIle-ArgV 544
Db     1626 TGTCCCCAGCGCACCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCCAGGG 1685

Qy      544 alAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleG 564
Db     1686 TGGCATGGCTGCCCCTGCCCTCCAGCATGAGCAATGGACAGGTGCTGAAGTACAAGATAG 1745

Qy      564 luTyrGlyLeuGlyLys---GluAspGlnValPheSerThrGluValProGlyAsnGluT 583
Db     1746 AGTACGGTTTGGGGAAGGAAGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAAA 1805

Qy      583 hrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaG 603
Db     1806 CACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTG 1865

Qy      603 lyThrGlyAla-GlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyVal 622
Db     1866 GCACTGGCGCTGGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTG 1925

Qy      623 HisAsnGlnSerHisValProPheAlaProAlaGluLeuLys------------------ 636
Db     1926 GACAACCAGAGCCATGTT------------ATGGAACTAAAAGGGGCAGAATGGAAAGGG 1973

Qy      637 ---------------ValArgAlaLysMetGluSerLeuValValSer------TrpGln 649
Db     1974 GGGACAGGTGCTCTGGTGGGCTTAGATATTTGGCCCCTTTTTTGGCTGGTTATTGGCAA 2033

Qy      650 ProProPro 652
Db     2034 CCCCCCCCC 2042
```

RESULT 2
AK083540

```
LOCUS       AK083540               4185 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:D030042G07 product:hypothetical Fibronectin
            type III domain containing protein, full insert sequence.
ACCESSION   AK083540
VERSION     AK083540.1  GI:26350620
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
```

```
  PUBMED   11076861
REFERENCE  4
  AUTHORS  Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
           Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
           Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
           Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
           Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
           Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
           Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
           Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
           Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
           Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
           Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
           Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
           Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
           Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
           Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
           Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
           Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
           Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
           and Hayashizaki,Y.
  TITLE    Functional annotation of a full-length mouse cDNA collection
  JOURNAL  Nature 409 (6821), 685-690 (2001)
  MEDLINE  21085660
   PUBMED  11217851
REFERENCE  5
  AUTHORS  The FANTOM Consortium and the RIKEN Genome Exploration Research
           Group Phase I & II Team.
  TITLE    Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
  JOURNAL  Nature 420, 563-573 (2002)
REFERENCE  6  (bases 1 to 4185)
  AUTHORS  Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
           Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
           Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
           Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
           Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
           Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
           Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
           Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
  TITLE    Direct Submission
  JOURNAL  Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT    cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site for further details.
           URL:http://genome.gsc.riken.go.jp/
           URL:http://fantom.gsc.riken.go.jp/.
FEATURES            Location/Qualifiers
     source         1..4185
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="FANTOM_DB:D030042G07"
                    /db_xref="taxon:10090"
                    /clone="D030042G07"
                    /tissue_type="whole body"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="9 days embryo"
     CDS            136..3195
                    /note="unnamed protein product; hypothetical Fibronectin
                    type III domain containing protein (InterPro|IPR003961,
                    evidence: InterPro)
                    putative"
                    /codon_start=1
                    /protein_id="BAC38947.1"
                    /db_xref="GI:26350621"
                    /translation="MDRVTALPSGVLQIYDVGPEDAGNYRCVAATIAHKRKSMEASLT
                    IVAANENRSFYMPTIIASPQNVTASLHQTVVLECMATGYPRPIISWSRLDHKSIDVFN
                    TRVLGNGNLIISDVKLQHAGVYVCRATTPGTRNFTVAMATLTVLAPPSFVEWPESLTR
                    PRAGTARFVCQAEGIPSPKMSWLKNGRRIHSNGRIKMYNSKLVINQIIPEDDAIYQCM
                    AENSQGSVLSRARLTVVMSEDRPSAPYNVHAETMSSSAILLAWERPLYNSDKVIAYSV
                    HYMKAEGLNNEEYQVVLGNDTTHYIIDDLEPDSNYTFYIVAYMPLGASQMSDHVTQNT
                    LEDVPLRPPEISLTSRSPTDILVSWLPIPAKYRRGQVVLYRLSFRLSTENAIQVVELP
                    GTVHEYLLEGLKPDSVYLVRITAATRVGLGESSVWTSHRTPKATSVKAPKSPELHLEP
                    LNCTTISVRWLQDTEDPAAIRGYKLFYKEEGQQEHGPIFLDTGDLLYTLSGLDPRRKY
                    HVRLLAYNNMEEGYQADQTVSTPGCVSVRDRMVPPPPPPHHLYAKANTSSSIFLHWRR
                    PAFTTAQVINYTIRCNPVGLQNASLVLYLQTSETHMLVQGLEPNTKYEFAVRLHVDQL
                    SSPWSPVVYHSTLPEAPTGPPVGVKVTLIEDDTALVSWKPPDGPETVVTRYTILYASR
                    KAWIAGEWQVLHREGAITMALLENLVAGNVYIVKISASNEVGEGPFSNSVELAVLPKD
                    ASESNQRPKRLDSSNAKVYSGYYHLDQKSMTGIAVGVGIALTCILICVLILIYRSKAR
                    KSSASKTAQSGTQPLSQASASVAAGSDMGKNLERATETAESLVPMMPSSFIDAKGGTD
                    LIINSYGPIIKNNTKKKWLFFQDTKKIKVEQTQRRFTQAVCFYQPGTTVLISDEDSPG
                    SPGQTASFPRPFGATALDTEHSANSEGSHETGDSGRFSHESNDEIHLSSVISSTPPTS
                    NPLAGGDSDGDAAPKKHGDPAQPLPAEQTSAPQTSAGLRYAAEGFPV"
BASE COUNT     1088 a   1049 c   1029 g   1019 t
ORIGIN

Alignment Scores:
Pred. No.:              3.26e-98      Length:        4185
Score:                  1610.50       Matches:       357
Percent Similarity:     55.37%        Conservative:  153
Best Local Similarity:  38.76%        Mismatches:    344
Query Match:            24.39%        Indels:        67
DB:                     11            Gaps:          15

US-09-754-997A-2 (1-1252) x AK083540 (1-4185)

Qy        144 PheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCys 163
              |||   :::||||||  |||        |||   |||   |||   ||||||||   |||
Db         10 TTTGAAGTCCATCCAGTGTCTACTGAGGTCCATGAAGGAGGAGTTGCTAGATTTTCATGC 69

Qy        164 HisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro 183
                          |||         :::||||||||||||   ::::::       :::|||
Db         70 AAGATTTCATCAACGCCGCCTGCAGTCATAACATGGGAATTTAATAGAACAGCCCTGCCT 129

Qy        184 GluGlu---ProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGln 202
                       |||:::         ||||||   :::|||||||||    ||||||
Db        130 ACAACCATGGACAGGGTGACTGCCCTGCCATCAGGAGTGTTGCAGATCTACGACGTTGGC 189

Qy        203 AspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSer 222
                    ||||||||:::|||||||||||||||               |||    :::|||   |||
Db        190 CCGGAGGATGCTGGGAATTACCGCTGTGTTGCAGCTACTATAGCTCATAAGAGGAAAAGC 249

Qy        223 GlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAsp 242
                 ||||||||||||||:::                    :::   |||          :::
Db        250 ATGGAGGCTTCGCTAACCATA---------------GTTGCAGCTAATGAGAACAGATCC 294

Qy        243 Val---------ValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsn 259
                          |||:::|||:::|||:::|||   |||                |||
Db        295 TTCTACATGCCAACCATTATAGCCAGTCCACAGAATGTAACCGCATCTCTACATCAGACA 354

Qy        260 ValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGln 279
              ||||||:::||||||:::|||:::         |||    |||      :::||||||   |||
Db        355 GTTGTGCTAGAGTGTATGGCCACAGGATATCCCAGACCCATCATTTCCTGGAGCCGTCTT 414

Qy        280 AspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeuLeuIle 298
              |||   |||   |||          :::         ||||||||       ||||||:::|||
Db        415 GATCACAAGTCCATTGATGTCTTCAACACTCGGGTACTTGGAAATGGCAACCTCATTATA 474

Qy        299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
              :::      :::   :::|||:::||||||||||||||||||||       |||    |||
Db        475 TCAGATGTCAAGCTGCAGCATGCTGGAGTGTATGTTTGTCGGGCCACCACCCCAGGCACA 534
```

```
Qy    319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGln 338
Db    535 CGTAACTTCACAGTTGCTATGGCAACGTTAACTGTATTAGCTCCTCCTTCATTTGTCGAA 594

Qy    339 AlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSer 358
Db    595 TGGCCTGAAAGTTTAACAAGGCCACGAGCTGGCACTGCACGATTTGTGTGCCAAGCAGAA 654

Qy    359 GlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGly 378
Db    655 GGAATTCCCTCACCCAAAATGTCCTGGTTGAAAAATGGGAGAAGGATACATTCAAATGGC 714

Qy    379 ArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAsp 398
Db    715 AGGATTAAAATGTAC------AACAGCAAATTGGTCATTAATCAGATTATCCCGGAAGAC 768

Qy    399 AlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaPro 418
Db    769 GATGCTATCTATCAGTGCATGGCTGAAAACAGCCAAGGATCTGTTCTGTCTCGAGCCAGA 828

Qy    419 LeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrPro 438
Db    829 CTGACTGTGGTAATGTCGGAAGATAGACCCAGCGCTCCTTACAACGTTCATGCTGAGACC 888

Qy    439 LeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIle 458
Db    889 ATGTCGAGCTCTGCTATCCTTCTGGCTTGGGAGAGACCACTGTACAATTCAGACAAGGTC 948

Qy    459 IleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPhe 478
Db    949 ATCGCCTACTCTGTACACTACATGAAGGCGGAAGGTTTAAATAACGAGGAATATCAGGTG 1008

Qy    479 AlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyr 498
Db   1009 GTCCTTGGAAACGACACGACTCATTACATCATTGATGACTTAGAACCCGACAGCAACTAT 1068

Qy    499 GluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeu 518
Db   1069 ACTTTCTACATTGTGGCGTACATGCCGCTGGGGGCCAGCCAGATGTCGGATCATGTGACA 1128

Qy    519 ValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsn 538
Db   1129 CAGAACACGCTAGAAGATGTGCCCCTAAGACCTCCCGAAATTAGCTTGACGAGTCGAAGT 1188

Qy    539 ProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnVal 558
Db   1189 CCCACCGACATTCTCGTCTCCTGGCTGCCAATCCCGGCCAAGTACCGGCGGGGCCAAGTG 1248

Qy    559 LeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluVal 578
Db   1249 GTGCTGTACCGCCTGTCCTTCCGCCTGAGCACTGAGAACGCCATCCAAGTTGTAGAACTC 1308

Qy    579 ProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgVal 598
Db   1309 CCGGGGACCGTGCACGAGTACCTTCTGGAAGGCCTGAAGCCAGACAGTGTCTATCTGGTT 1368

Qy    599 ArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArg 618
Db   1369 CGGATTACTGCAGCTACCCGTGTGGGGCTGGGAGAGTCGTCGGTGTGGACCTCACACAGG 1428

Qy    619 ThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArg 638
Db   1429 ACGCCCAAGGCT---ACGAGCGTGAAAGCCCCGAAGTCTCCT---GAACTGCATTTGGAA 1482

Qy    639 ---AlaLysMetGluSerLeuValValSerTrp---GlnProProProHisProThrGln 656
Db   1483 CCACTGAATTGTACTACCATCTCTGTGAGGTGGCTGCAAGATACAGAGGACCCGGCCGCT 1542

Qy    657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
Db   1543 ATTCGGGGCTACAAGCTGTTTTATAAAGAGGAAGGGCAGCAAGAACAC------------ 1590

Qy    677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
Db   1591 ---------------------------------------GGACCTATCTTCCTAGACACC 1611

Qy    697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
Db   1612 GGTGACCTGCTGTATACGCTCAGTGGCTTGGACCCCAGAAGAAAATACCACGTGAGGCTC 1671

Qy    717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
Db   1672 CTGGCTTACAACAACATGGAAGAAGGCTACCAGGCC---------------GACCAGACT 1716

Qy    737 AlaProThrProAspLeuProIleGlnArgGly------------ProProLeuProPro 752
Db   1717 GTCAGCACTCCCGGATGCGTGTCCGTCCGTGATCGCATGGTCCCTCCACCACCGCCACCC 1776

Qy    753 AlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysPro 772
Db   1777 CACCATCTCTACGCGAAGGCGAATACCTCCTCTTCCATCTTCCTGCACTGGAGGAGGCCT 1836

Qy    773 AspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArg 792
Db   1837 GCGTTCACCACCGCCCAAGTCATTAACTACACCATCCGCTGTAACCCTGTTGGCCTGCAG 1896

Qy    793 AsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeu 812
Db   1897 AATGCCTCTCTGGTTCTCTACCTGCAAACATCAGAAACCCACATGCTGGTCCAAGGACTA 1956

Qy    813 LysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyPro 832
Db   1957 GAACCAAACACCAAGTATGAATTTGCTGTCCGACTGCATGTGGATCAGCTGTCCAGTCCC 2016

Qy    833 PheGlySerValValGluArgSerThrLeuProAspArgProSerThrProProSerAsp 852
Db   2017 TGGAGTCCTGTGGTCTACCATTCTACATTGCCAGAAGCACCCACAGGCCCCCCGGTAGGG 2076

Qy    853 LeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGlu 872
Db   2077 GTAAAGGTGACCCTGATAGAAGACGACACTGCCCTTGTTTCCTGGAAACCCCCAGATGGC 2136

Qy    873 ProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGlu 892
Db   2137 CCAGAGACAGTGGTGACACGCTACACCATCTTGTATGCTTCCAGGAAGGCCTGGATTGCG 2196

Qy    893 HisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeu 912
Db   2197 GGAGAGTGGCAGGTTCTTCACCGAGAAGGGGCAATCACCATGGCTTTGCTGGAAAACCTG 2256

Qy    913 GluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyPro 932
Db   2257 GTGGCAGGGAACGTGTACATTGTCAAGATATCTGCGTCCAACGAGGTGGGAGAAGGGCCC 2316

Qy    933 PheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--------- 949
Db   2317 TTTTCAAACTCCGTGGAGCTGGCCGTCCTCCCCAAGGATGCTTCAGAGTCGAACCAGAGG 2376

Qy    950 ---------------------------------------------------LeuAspVal 952
Db   2377 CCCAAGCGTTTGGATTCTTCCAATGCCAAAGTTTATTCCGGCTATTACCATCTGGACCAA 2436

Qy    953 HisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCys 972
Db   2437 AAGTCAATGACGGGCATCGCAGTAGGCGTTGGCATAGCCTTGACCTGTATCCTCATCTGT 2496

Qy    973 MetCysAlaGlyLeu---ArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSer 991
Db   2497 GTTCTCATCTTGATCTACCGAAGCAAAGCCAGGAAGTCATCCGCCTCTAAGACAGCGCAG 2556

Qy    992 SerGlyThr------------ProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGly 1007
Db   2557 AGTGGAACCCAGCCGTTATCCCAAGCCAGTGCCTCTGTAGCAGCAGGGAGTGACATGGGG 2616

Qy   1008 ProProSerValProAlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAsp 1027
```

```
Db        2617 AAGAACCTGGAGAGAGCTACAGAGACTGCAGAGTCCTTGGTACCCATGATGCCCAGCAGC 2676

Qy        1028 Trp 1028
               :::
Db        2677 TTC 2679

RESULT 3
BI185886
LOCUS       BI185886                 900 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  UNL-P-FN-df-h-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-df-h-02-0-UNL 3', mRNA sequence.
ACCESSION   BI185886
VERSION     BI185886.1  GI:14660295
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Mamm. Genome 14 (1), 65-70 (2003)
  MEDLINE   22419904
   PUBMED   12532269
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..900
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-df-h-02-0-UNL"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UNL-P-FN"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      149 a    304 c    291 g    150 t      6 others
ORIGIN

Alignment Scores:
Pred. No.:              8.27e-80       Length:         900
Score:                  1326.00        Matches:        262
Percent Similarity:     97.12%         Conservative:   8
Best Local Similarity:  94.24%         Mismatches:     8
Query Match:            20.08%         Indels:         1
DB:                     12             Gaps:           0

US-09-754-997A-2 (1-1252) x BI185886 (1-900)

Qy         247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
               ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db           7 GCTGCCCCGGAGAACACCACAGTGGTGTCCGGCCAGAGCGTGGTGATGGAGTGTGTGGCC 66

Qy         267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
               ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db          67 TCGGCTGACCCCACCCCTATCGTGTCCTGGGTCCGACAGGACGGGAAGCCTATTTCTACC 126

Qy         287 AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSer 306
               |||||||||||||||||||||||||||||||||||||   ||||||   |||   |||||
Db         127 GATGTCATCGTCCTGGGCCGCACCAACCTGCTAATCACCAGCGCGCANCCCGG-CACTCT 185

Qy         307 GlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAla 326
               ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db         186 GGCGTCTACGTCTGCAGAGCCAACAAGCCCCGTACGCGCGACTTTGCCACCGCCGCCGCC 245

Qy         327 GluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThr 346
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         246 GAGCTCCGCGTGCTGGCGGCTCCTGCCATCTCGCAGGCGCCCGAGGCGCTATCTCGGACG 305

Qy         347 ArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHis 366
               ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         306 AGGGCGAGCACCGCGCGCTTCGTGTGCCGCGCAACGGGTGAGCCGCGGCCCGCGTTGCGG 365

Qy         367 TrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGlyGly 386
               |||||||||:::|||   ||||||||||||||||||||||||||||||||||||||||||
Db         366 TGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGTGTCAAGGTGCAGGGCGGCGGC 425

Qy         387 GlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAla 406
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         426 AGCAGCCTGGTCATCACGCAGATCGGCCTGCAGGATGCCGGCTACTACCAGTGCGTGGCC 485

Qy         407 GluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGly 426
               ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         486 GAGAACAACGCGGGCACGGCGTGCGCCGCCGCGCCGCTGGCAGTAGTAGTGCGCGAGGGT 545

Qy         427 LeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeuVal 446
               ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db         546 CTGCCCAGCGCCCCCACGCGGGTCACCGCCACGCCCCTGAGCAGCTCCACTGTGCTGGTG 605

Qy         447 AlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGln 466
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         606 GCCTGGGAGCGGCCCGAGCTGCACAGCGAGCAGATCATCGGTTTCTCCCTTCACTACCAG 665

Qy         467 LysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGlu 486
               ||||||||||||:::|||:::|||||||||||||||||||||||||||||||||||||||
Db         666 AAGGCACGGGGCATGGACAGTGTGGAATACCAGTTTGCCGTGAACAATGACACCACTGAG 725

Qy         487 LeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSer 506
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         726 CTACAGGTTCGGGACCTGGAGCCCAACACAGACTATGAGTTCTATGTGGTGGCCTACTCA 785

Qy         507 GlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAsp 524
               ||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db         786 CAGCTGGGGGCCAGCCGCACCTCCACCCCAGCGCTGGTCCACACATTGGATGAT 839

RESULT 4
CA317106
LOCUS       CA317106                 817 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FW0-cbm-d-21-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
            IMAGE: 6811462 5', mRNA sequence.
ACCESSION   CA317106
VERSION     CA317106.1  GI:24535230
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
                 Tissue Procurement: Dr. Jim Lin, University of Iowa
                  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                  Clone Distribution: MGC clone distribution information can be
                 found through the I.M.A.G.E. Consortium/LLNL at:
                 http://image.llnl.gov
                  This clone was contributed by the Brain Molecular Anatomy Project
                 (BMAP)
                 Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..817
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6811462"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FW0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT     180 a    254 c    234 g    148 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               1.46e-77      Length:         817
Score:                   1292.00       Matches:        245
Percent Similarity:      98.39%        Conservative:   0
Best Local Similarity:   98.39%        Mismatches:     4
Query Match:             19.56%        Indels:         1
DB:                      14            Gaps:           0

US-09-754-997A-2 (1-1252) x CA317106 (1-817)

Qy       1004 AlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisPro 1023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCTCGACTTGGGCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCT 61

Qy       1024 ArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSer 1043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGC 121

Qy       1044 LeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIleSerTrpAla 1063
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCT 181

Qy       1064 GlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySer 1083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGT 241

Qy       1084 GlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeu 1103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTG 301

Qy       1104 LeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLysProSerPro 1123
              ||||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db        302 CTGCTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCA 361

Qy       1124 AlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGlyAlaSerLys 1143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAA 421

Qy       1144 GlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThrAlaGluThr 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACT 481

Qy       1164 LeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGlu 1183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTGGGCAGGGAG 541

Qy       1184 LeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeuProGluAla 1203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 CTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTGCCAGAAGCA 601

Qy       1204 AlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGluGluAlaPro 1223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAGGAGGCCCCT 661

Qy       1224 GlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSerLeuProArg 1243
              ||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        662 GGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGCTTCCCAGN- 720

Qy       1244 AlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||    |||
Db        721 GCCCCTGTCTCCTCTGCTCAAGGTCCC 747

RESULT 5
BU701299
LOCUS       BU701299                 740 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-EX0-bwy-i-18-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
            IMAGE: 5704793 5', mRNA sequence.
ACCESSION   BU701299
VERSION     BU701299.1  GI:23624921
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..740
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 5704793"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EX0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
```

```
              cDNA was size selected according to mRNA size fraction,
              ligated with EcoR I adaptor, digested with Not I, and then
              cloned directionally into pYX-Asc vector. The library tag
              sequence located between the Not I site and the polyA tail
              , is GTGCGTGGAA. This library was created for the
              University of Iowa Mouse Brain Molecular Anatomy Project
              (BMAP): 'Gene Discovery in the Developing Mouse Nervous
              System', supported by National Instutes of Mental Health
              (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      173 a    225 c    200 g    142 t
ORIGIN

Alignment Scores:
Pred. No.:              6.52e-75      Length:         740
Score:                  1252.00       Matches:        242
Percent Similarity:     97.98%        Conservative:   0
Best Local Similarity:  97.98%        Mismatches:     5
Query Match:            18.96%        Indels:         3
DB:                     13            Gaps:           0

US-09-754-997A-2 (1-1252) x BU701299 (1-740)

Qy      654 ProThrGlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAla 673
            ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db        2 CCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCA 61

Qy      674 AspGlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArg 693
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62 GATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGG 121

Qy      694 LeuLysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGlu 713
            ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db      122 CTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAG 181

Qy      714 ValLysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLys 733
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      182 GTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAG 241

Qy      734 ThrGluLysAlaProThrProAspLeuProIleGlnArgGlyProProLeuProProAla 753
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      242 ACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCC 301

Qy      754 HisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAsp 773
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      302 CATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGAC 361

Qy      774 PheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsn 793
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      362 TTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAAT 421

Qy      794 AlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLys 813
            ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db      422 GCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGA-ATTCTCATTGGCGGCCTGAAA 480

Qy      814 ProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPhe 833
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTT 540

Qy      834 GlySerValValGluArgSerThrLeuProAspArgProSerThrProProSerAspLeu 853
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTG 600

Qy      854 ArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGluPro 873
            ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db      601 CGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCACG-GAGCCC 659

Qy      874 AsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHis 893
            |||||||||||||||||||||||||||||||||||||||    |||||||||||||||||
Db     -660 AATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAA-CACACCCAGCCCGAACAC 718

Qy      894 GlnTrpThrLeuLeuThrThr 900
            |||||||||||||||||||||
Db      719 CAGTGGACACTGCTCACCACA 739
```

```
RESULT 6
CD351162
LOCUS       CD351162                 699 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-FY0-cft-p-13-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6852446 5', mRNA sequence.
ACCESSION   CD351162
VERSION     CD351162.1  GI:31142737
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..699
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6852446"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      135 a    221 c    202 g    139 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              1.65e-72      Length:         699
Score:                  1216.00       Matches:        227
Percent Similarity:     98.71%        Conservative:   2
Best Local Similarity:  97.84%        Mismatches:     3
Query Match:            18.41%        Indels:         0
DB:                     14            Gaps:           0

US-09-754-997A-2 (1-1252) x CD351162 (1-699)

Qy      919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3 TTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGAT 62

Qy      939 ValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIle 958
            ||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||
```

```
Db         63 GTGATTACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATC 122

Qy        959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 ATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGA 182

Qy        979 GlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnPro 998
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCA 242

Qy        999 AlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGlu 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAG 302

Qy       1019 SerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLys 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAG 362

Qy       1039 AlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArg 1058
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGA 422

Qy       1059 LysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGlu 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 AAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAG 482

Qy       1079 LeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGly 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGA 542

Qy       1099 ThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArg 1118
              |||||||||||||||||||||||||||||||||||||||||  |||  ||||||||||||
Db        543 ACCGGGCAGACACTGCTGCTGCAAGCCCTGGTATATGACGCCATANAGAGCAACGGGAGA 602

Qy       1119 LysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAsp 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 AAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGAC 662

Qy       1139 PheGlyAlaSerLysGlyCysProAspLeuHisLeu 1150
              |||||||||||   ||||||||||||||||||||||
Db        663 TTCGGTGCATCCANAGGATGTCCTGACCTCCACCTC 698
```

```
RESULT 7
CB524810
LOCUS       CB524810                 713 bp    mRNA    linear   EST 28-MAR-2003
DEFINITION  UI-M-FY0-cey-g-12-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6843781 5', mRNA sequence.
ACCESSION   CB524810
VERSION     CB524810.1  GI:29358165
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..713
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6843781"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      141 a    235 c    185 g    148 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:              1.09e-71        Length:         713
Score:                  1204.00         Matches:        230
Percent Similarity:     97.89%          Conservative:   2
Best Local Similarity:  97.05%          Mismatches:     5
Query Match:            18.23%          Indels:         1
DB:                     14              Gaps:           0

US-09-754-997A-2 (1-1252) x CB524810 (1-713)

Qy        831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db          3 GGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTC-ACACCTCCT 61

Qy        851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCC 121

Qy        871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAG 181

Qy        891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCAT 241

Qy        911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCT 301

Qy        931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLeu 950
              ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        302 GGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAAAGACATTCTCAGACTCCTTG 361

Qy        951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GATGTGCACGCCGTCACGGGCATCATCGTGGGTGTCTGCCTGGGCCTTCTCTGCCTCCTG 421

Qy        971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSer 990
              |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        422 GCCTGCATGTGTGCTGGCCTACGACGAAGCTCCCACAGGGAAGCCCTGCCCGGATTGTCC 481

Qy        991 SerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSer 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 TCCTCAGGCACCCCAGGAAACCCAGCGCTCTACACAAGAGCTCGACTTGGGCCCCCCAGT 541

Qy       1011 ValProAlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAspTrpSerPro 1030
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
```

```
Db        542 GTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGTCNCCAGGATTGGTCCCCA 601

Qy       1031 ProProSerAspValGluAspLysAlaGluValHisSerLeuMetGlyGlySerValSer 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 CCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCA 661

Qy       1051 AspCysArgGlyHisSerLysArgLysIleSerTrpAlaGlnAlaGlyGly 1067
              |||||||||||||||||| | ||||||||||||   |||    ||||||||
Db        662 GATTGCCGGGGCCACTCCANGAGAAAGATCTCCTGNGCTCANGCAGGGGGA 712

RESULT 8
CD352400
LOCUS       CD352400                 741 bp    mRNA    linear   EST 29-MAY-2003
DEFINITION  UI-M-GI0-cgj-p-07-0-UI.r1 NIH_BMAP_GI0 Mus musculus cDNA clone
            IMAGE: 6856280 5', mRNA sequence.
ACCESSION   CD352400
VERSION     CD352400.1  GI:31143911
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..741
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6856280"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GI0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector . The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      139 a    238 c    219 g    145 t
ORIGIN

Alignment Scores:
Pred. No.:               1.56e-71       Length:         741
Score:                   1202.00        Matches:        240
Percent Similarity:      98.37%         Conservative:   2
Best Local Similarity:   97.56%         Mismatches:     4
Query Match:             18.20%         Indels:         0
DB:                      14             Gaps:           0

US-09-754-997A-2 (1-1252) x CD352400 (1-741)

Qy        106 GluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerPro 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTGGCCCACAGCCCG 61

Qy        126 LeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSer 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTCGAAGACTTCTCT 121

Qy        146 LeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThr 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTTGAATGCCACACC 181

Qy        166 LysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnValThrValProGluGlu 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 AAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACCGTGCCTGAGGAG 241

Qy        186 ProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAsp 205
                 |||||||||||||||||       :::|||||||||||||||||||||||||||
Db        242 TCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATGTCCAGGACAGTGAT 301

Qy        206 AlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAla 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTCAGCCAGGAGGCC 361

Qy        226 SerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIle 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTCATT 421

Qy        246 ValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysVal 265
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        422 GTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTGATGGAGTGCGTG 481

Qy        266 AlaSerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSer 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 GCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATCTCC 541

Qy        286 ThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCAC 601

Qy        306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db        602 TCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGCT 661

Qy        326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGG 721

Qy        346 ThrArgAlaSerThrAla 351
              ||||||||||||||||||
Db        722 ACGCGGGCCAGCACCGCG 739

RESULT 9
CA751384
LOCUS       CA751384                 719 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  UI-M-FO0-cdn-j-21-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE: 6830806 5', mRNA sequence.
ACCESSION   CA751384
VERSION     CA751384.1  GI:25579257
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
          Tissue Procurement: Dr. Jim Lin, University of Iowa
           cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
           cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
           Clone Distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
           This clone was contributed by the Brain Molecular Anatomy Project
          (BMAP)
          Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6830806"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      112 a    234 c    235 g    135 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              1.18e-69      Length:          719
Score:                  1174.00       Matches:         234
Percent Similarity:     98.33%        Conservative:    1
Best Local Similarity:  97.91%        Mismatches:      4
Query Match:            17.78%        Indels:          0
DB:                     14            Gaps:            0

US-09-754-997A-2 (1-1252) x CA751384 (1-719)

Qy        225 AlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValVal 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAGGATGTGGTC 60

Qy        245 IleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCys 264
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db         61 ATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTGATGGAGTGC 120

Qy        265 ValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIle 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGAAAGCCTATC 180

Qy        285 SerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArg 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCGCAGCCTCGG 240

Qy        305 HisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAla 324
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        241 CACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTCGCCACTGCG 300

Qy        325 AlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSer 344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCG 360

Qy        345 ArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAla 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCG 420

Qy        365 LeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGly 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGCACTGGCTGCATGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGC 480

Qy        385 GlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCys 404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGC 540

Qy        405 ValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArg 424
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        541 GTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTANTGGTGCGC 600

Qy        425 GluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerVal 444
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAGGGGCTGCCCANCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTG 660

Qy        445 LeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeu 463
              |||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTGGTGGCCTGNGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTT 717

RESULT 10
CB518504
LOCUS       CB518504                631 bp    mRNA    linear   EST 28-MAR-2003
DEFINITION  UI-M-GH0-cec-h-04-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
            IMAGE: 6837653 5', mRNA sequence.
ACCESSION   CB518504
VERSION     CB518504.1  GI:29351859
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6837653"
                     /tissue_type="Whole brain"
                     /dev_stage="1, 5, and 15 days newborn"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GH0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CGAACTGAAT. This library was created for the University
```

```
              Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
              Developing Mouse Nervous System', supported by National
              Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
              program coordinator."
BASE COUNT      151 a    183 c    180 g    117 t
ORIGIN

Alignment Scores:
Pred. No.:              1.37e-64      Length:         631
Score:                  1098.00       Matches:        208
Percent Similarity:     99.52%        Conservative:   0
Best Local Similarity:  99.52%        Mismatches:     1
Query Match:            16.63%        Indels:         0
DB:                     14            Gaps:           0

US-09-754-997A-2 (1-1252) x CB518504 (1-631)

Qy        470 GlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnVal 489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTT 62

Qy        490 ArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGly 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGGGACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGG 122

Qy        510 AlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspValProSerAlaAla 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCA 182

Qy        530 ProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeuProLeu 549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTG 242

Qy        550 ProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLys 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAG 302

Qy        570 GluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSer 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCA 362

Qy        590 LeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGly 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGA 422

Qy        610 ValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSerHisValPro 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCC 482

Qy        630 PheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuValValSerTrpGln 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAG 542

Qy        650 ProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThr 669
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        543 CCGCCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACA 602

Qy        670 GluGluGluAlaAspGlyAspArgPro 678
              |||||||||||||||||||||||||||
Db        603 GAGGAGGAGGCAGATGGTGACCGCCCC 629

RESULT 11
CB988160
LOCUS       CB988160                  759 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  AGENCOURT_13892946 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30346067 5', mRNA sequence.
ACCESSION   CB988160
VERSION     CB988160.1  GI:30282680
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
             cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM384  row: p  column: 12
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1..759
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30346067"
                     /tissue_type="Human Placenta"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_147"
                     /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                     alI-XhoI; Site_2: BamH; Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected  for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using  the Cap-trapper method (Carninci, in
                     preparation). Library constructed  by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: This is
                     a NIH_MGC library."
BASE COUNT      161 a    229 c    237 g    131 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              2.02e-64      Length:         759
Score:                  1097.00       Matches:        218
Percent Similarity:     94.17%        Conservative:   8
Best Local Similarity:  90.83%        Mismatches:     11
Query Match:            16.61%        Indels:         3
DB:                     14            Gaps:           1

US-09-754-997A-2 (1-1252) x CB988160 (1-759)

Qy        378 GlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGln 397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 GGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAG 86

Qy        398 AspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAla 417
              ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||
Db         87 GACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCG 146

Qy        418 ProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThr 437
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 TCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACG 206

Qy        438 ProLeuSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGln 457
              |||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||||
Db        207 CCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAG 266

Qy        458 IleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGln 477
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        267 ATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAG 326

Qy        478 PheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAsp 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 TTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGAT 386

Qy        498 TyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAla 517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
```

```
Db        387 TATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCA 446

Qy        518 LeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerPro 537
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||
Db        447 CTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCC 506

Qy        538 AsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGln 557
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        507 AACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAG 566

Qy        558 ValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGlu 577
              |||:::|||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        567 GTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAG 626

Qy        578 ValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArg 597
              |||    |||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        627 GTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGA 686

Qy        598 ValArgIle-SerAlaGlyThrGly------AlaGlyTyrGlyValProSerGlnTrp 614
              |||||||| ||||||||||||            |||        |||   :::|||
Db        687 GTACGGATTTTCGGCTGGTACAGCAGCCCGGGCTTCNGGGGCCCCCCCTCCCAGGTGG 744
```

```
RESULT 12
BY712424
LOCUS       BY712424                703 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY712424 RIKEN full-length enriched, 10, 11 days embryo whole body
            Mus musculus cDNA clone 2810488H22 5', mRNA sequence.
ACCESSION   BY712424
VERSION     BY712424.1  GI:27123733
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..703
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2810488H22"
                     /tissue_type="whole body"
                     /dev_stage="10, 11 days embryo"
                     /clone_lib="RIKEN full-length enriched, 10, 11 days embryo
                     whole body"
BASE COUNT      164 a    185 c    218 g    134 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              1.75e-63      Length:         703
Score:                  1082.50       Matches:        210
Percent Similarity:     86.53%        Conservative:   2
Best Local Similarity:  85.71%        Mismatches:     3
Query Match:            16.39%        Indels:         30
DB:                     14            Gaps:           1

US-09-754-997A-2 (1-1252) x BY712424 (1-703)

Qy        497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCA 61

Qy        517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGC 121

Qy        537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        122 CCCAACCCCTCGGACATGAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGA 181

Qy        557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACC 241

Qy        577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTAC 301

Qy        597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAG 361

Qy        617 HisArgThrProGlyValHisAsnGlnSerHis-ValProPheAlaProAlaGluLeuLy 636
              |||||||||||||||||||||||||||||||||| :::
Db        362 CACAGGACACCTGGTGTGCACAACCAGAGCCATGATC---------------------- 398

Qy        636 sValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGl 656

Db        398 ------------------------------------------------------------ 398

Qy        656 nIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAs 676
                  ||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        399 ----TCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGA 454

Qy        676 pArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLy 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 CCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAA 514

Qy        696 sLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLe 716
              |||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        515 GAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCT 574

Qy        716 uValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLy 736
              |||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        575 CGTAGCTTTCAACANACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAA 634

Qy        736 sAlaProThrPro 740
              |||||||||||||
Db        635 GGCGCCCACGCCA 647
```

RESULT 13
BU057120

```
LOCUS       BU057120                 591 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FO0-caa-e-21-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE:6408572 5', mRNA sequence.
ACCESSION   BU057120
VERSION     BU057120.1  GI:22497197
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6408572"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      126 a    179 c    179 g    106 t      1 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:               3.55e-61        Length:          591
Score:                   1047.00         Matches:         193
Percent Similarity:      98.98%          Conservative:    1
Best Local Similarity:   98.47%          Mismatches:      2
Query Match:             15.85%          Indels:          0
DB:                      13              Gaps:            0
```

US-09-754-997A-2 (1-1252) x BU057120 (1-591)

```
Qy       1005 ArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisProArg 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 CGACTTGGGCCCCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTCGTGCATCCTCGT 62

Qy       1025 ProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSerLeu 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAAGTACACAGCCTT 122

Qy       1045 MetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIleSerTrpAlaGln 1064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 ATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATCTCCTGGGCTCAG 182

Qy       1065 AlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySerGly 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCCCAGGGTAGTGGT 242

Qy       1085 ProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeu 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGGCAGACACTGCTG 302

Qy       1105 LeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLysProSerProAla 1124
              ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        303 CTGCAAGCCCTGGTATATGACGCCATAAAGAGCAACGGGAGAAAGAAGCCGTCCCCAGCC 362

Qy       1125 CysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGlyAlaSerLysGly 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGTGCATCCAAAGGA 422

Qy       1145 CysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThrAlaGluThrLeu 1164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACTGCAGAGACTCTG 482

Qy       1165 ProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGluLeu 1184
              ||||||||||||||||||||||||||||||||||||||||||   |||||||:::||||||
Db        483 CCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCANACTGGCTGGGCAAGGAGCTG 542

Qy       1185 GlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 590
```

RESULT 14
AK051027

```
LOCUS       AK051027                2906 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:D030056K15 product:putative neuronal cell
            adhesion molecule, full insert sequence.
ACCESSION   AK051027
VERSION     AK051027.1  GI:26341679
```

```
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2906)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2906
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D030056K15"
                     /db_xref="taxon:10090"
                     /clone="D030056K15"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9 days embryo"
     CDS             228..2669
                     /note="unnamed protein product; putative
                     putative neuronal cell adhesion molecule (MGD|MGI:1202390,
                     GB|NM_008988, evidence: BLASTN, 100%, match=726)"
                     /codon_start=1
                     /protein_id="BAC34502.1"
                     /db_xref="GI:26341680"
                     /translation="MAEPRTASPRRLPALRRPGFLPPLLPPPPPPLLLLLLLLPLPAP
                     SLGLGHSAELAFSVEPNDDIANPGQPIVLGCKVEGTPPVQVSWRKNGAELPEGTHTTL
                     LANGSLLIHHFRLEQGGSPSDEGDYECVAQNRFGLLVSRKARLQAATMSDFHVHPQAV
                     TGEEGGVARFQCQIHGLPKPLITWEKNRVPIDTDDERYTLLPKGVLQITGLRAEDSGI
                     FHCVASNIASVRVSHGARLTVSGSGSGTYKEPTILVGPENLTLTVHQTAVLECVATGN
                     PRPIVSWSRLDGRPIGVEGIQVLGTGNLIISDVTVQHSGVYVCAANRPGTRVRRTAQG
                     RLVVQAPAEFVQHPQSISRPAGTTAMFTCQAQGEPPPHVTWLKNGQVLGAGGHVRLKN
                     NNSTLSISGVGPEDEAIYQCVAENIAGSSQASARLTVLWAEGLPGPPRNVRAVSVSST
                     EVRVSWSEPLAHTKEIIGYVLHIRKAADSPKLEYQEAVSKSTFQHLVRDLEPSTAYSF
                     YIKAYTPRGASLASVPTLASTLGEAPVPPPLSVRLLGSSSLQLLWKPWPRLAQHNGGF
                     KLFYRPVSATSFTGPILLPGTVSSYNLSQLDPSTVYEVKLLAYNQHGDGNATVRFVSL
                     KGASERTALTPPCDCRKEDVTNHTSTTGIVIGIHIGVTCIIFCVLFLLFGQRGRVLLC
                     KDVENQLSPPQGPRSQRDPGILALNGLSRGEGGQLSRDEKPVDAKELEQLFPTAGSAA
                     QPGSTPTDPAAPAPCEETQLSMVQLQGFNLVAGRTTEATSPCAGPGPVPAPQDIGPVP
                     LSEGQTQPPAVAAPQ"
BASE COUNT      582 a    937 c    835 g    552 t
ORIGIN

Alignment Scores:
Pred. No.:              2.97e-60      Length:         2906
Score:                  1046.00       Matches:        328
Percent Similarity:     38.49%        Conservative:   135
Best Local Similarity:  27.27%        Mismatches:     307
Query Match:            15.84%        Indels:         434
DB:                     11            Gaps:           34

US-09-754-997A-2 (1-1252) x AK051027 (1-2906)

Qy          24 LeuProLeuProGlnGluThr---------------ThrValLysLeuSerCysAspGlu 38
               ||||||||||||      :::               :::   :::|||:::
Db         342 TTGCCGCTGCCTGCGCCAAGCCTGGGTCTGGGCCATTCTGCTGAACTGGCATTTTCTGTG 401

Qy          39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
                  |||         |||      |||   |||   :::||||||   |||   :::
Db         402 GAGCCAAATGATGATATTGCCAACCCTGGGCAGCCCATAGTGCTGGGCTGCAAAGTGGAG 461

Qy          59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
```

```
Db    462 GGCACC---------CCTCCAGTGCAAGTCTCCTGGAGGAAGAATGGAGCAGAGCTACCA 512
Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSer---ProLeu 97
Db    513 GAGGGCACCCACACCACTCTGCTGGCCAATGGCTCCTTGCTGATCCACCACTTCAGGCTG 572
Qy     98 GluGlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySer 117
Db    573 GAGCAAGGAGGTAGCCCTTCAGAT--------------------------GAGGGTGAC 605
Qy    118 TyrSerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLys 137
Db    606 TATGAATGTGTGGCTCAGAACCGCTTTGGGCTGCTGGTCAGTCGGAAGGCTCGTCTCCAG 665
Qy    138 LeuAlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGly 157
Db    666 GCTGCAACTATGTCCGACTTCCACGTACACCCCCAGGCCGTCACGGGTGAGGAGGGTGGT 725
Qy    158 ThrAlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLys 177
Db    726 GTGGCCCGCTTCCAGTGCCAAATCCATGGGCTTCCCAAGCCCCTGATCACATGGGAGAAG 785
Qy    178 AspGlnValThrVal---ProGluGluProArgLeuIleThrLeuProLysTrpLeuLeu 196
Db    786 AACAGAGTTCCCATTGACACGGACGATGAGAGGTACACACTACTGCCCAAGGGAGTCCTG 845
Qy    197 GlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSer 216
Db    846 CAGATAACAGGACTTCGGGCTGAGGACAGCGGCATCTTCCACTGTGTGGCCTCAAACATT 905
Qy    217 AlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGlu 236
Db    906 GCTAGTGTGCGGGTCAGCCACGGGGCCAGGCTCACTGTGTCA------GGCTCGGGCTCA 959
Qy    237 AlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSer 256
Db    960 GGGACCTACAAGGAGCCCACCATTCTTGTG---GGGCCTGAGAACCTCACCCTGACAGTG 1016
Qy    257 GlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrp 276
Db   1017 CACCAGACTGCGGTGCTGGAGTGTGTCGCCACGGGCAACCCGCGCCCCATTGTGTCCTGG 1076
Qy    277 ValArgGlnAspGlyLysProIleSerThrAspValIle---ValLeuGlyArgThrAsn 295
Db   1077 AGCCGCTTGGATGGCCGCCCCATTGGGGTGGAGGGCATCCAGGTGCTGGGCACAGGAAAC 1136
Qy    296 LeuLeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLys 315
Db   1137 CTTATCATCTCAGATGTGACTGTCCAGCATTCAGGTGTCTATGTCTGTGCAGCCAACAGG 1196
Qy    316 ProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAla 335
Db   1197 CCTGGCACCCGGGTGAGAAGGACGGCCCAGGTCGGCTGGTGGTACAAGCTCCAGCAGAG 1256
Qy    336 IleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCys 355
Db   1257 TTTGTTCAGCATCCCCAGTCCATCTCCAGGCCTGCTGGGACCACAGCCATGTTCACCTGC 1316
Qy    356 ArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArg 375
Db   1317 CAAGCGCAGGGGGAGCCACCGCCTCATGTCACATGGTTGAAGAATGGACAAGTGCTGGGG 1376
Qy    376 ProAsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGly 395
Db   1377 GCAGGAGGCCATGTCAGGCTCAAGAACAATAACAGCACACTGAGCATTTCTGGAGTCGGA 1436
Qy    396 LeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAla 415
Db   1437 CCTGAAGATGAGGCCATCTACCAATGTGTGGCTGAGAACATTGCTGGCTCGTCGCAAGCC 1496
Qy    416 AlaAlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThr 435
Db   1497 AGCGCCAGGCTGACCGTGCTGTGGGCTGAGGGGTTGCCAGGCCCCCCACGCAATGTCCGA 1556
Qy    436 AlaThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSer 455
Db   1557 GCTGTCTCTGTGTCCTCCACTGAAGTTCGAGTGTCCTGGAGTGAGCCTCTGGCCCACACC 1616
Qy    456 GluGlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGlu 475
Db   1617 AAAGAGATCATTGGCTATGTCTTGCACATCAGGAAGGCTGCCGACTCACCCAAGCTAGAG 1676
Qy    476 TyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsn 495
Db   1677 TACCAGGAAGCTGTCAGCAAGAGCACCTTTCAGCACCTGGTGAGAGACCTTGAGCCCTCC 1736
Qy    496 ThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSer 515
Db   1737 ACTGCCTACAGCTTCTACATCAAAGCCTACACACCACGAGGGGCCAGTTTAGCCTCTGTC 1796
Qy    516 ProAlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSer 535
Db   1797 CCCACCCTGGCCAGCACCCTGGGTGAAGCCCCT---GTCCCACCCCCATTATCTGTGCGG 1853
Qy    536 SerProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsn 555
Db   1854 TTGCTGGGCAGCTCCTCTCTACAACTGCTATGGAAGCCCTGGCCA--------------- 1898
Qy    556 GlyGlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSer 575
Db   1898 ------------------------------------------------------------ 1898
Qy    576 ThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysVal 595
Db   1898 ------------------------------------------------------------ 1898
Qy    596 TyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMet 615
Db   1899 ---------CGGCTGGCCCAA--------------------------------------- 1910
Qy    616 GlnHisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeu 635
Db   1911 --------------------CACAAT---------------------------------- 1916
Qy    636 LysValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
Db   1916 ------------------------------------------------------------ 1916
Qy    656 GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGly 675
Db   1917 ------GGAGGCTTCAAGCTGTTTACCGCCCAGTCAGTGCAACCTCC------------ 1958
Qy    676 AspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLys 695
Db   1959 ------------------------------------TTCACTGGCCCCATCTTGCTCCCT 1982
Qy    696 LysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLys 715
Db   1983 GGGACTGTCTCCTCTTACAACCTCAGCCAGCTTGACCCCAGCACCGTGTACGAGGTGAAG 2042
Qy    716 LeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGlu 735
Db   2043 CTCCTGGCCTACAACCAGCATGGAGATGGA------------------------------ 2072
Qy    736 LysAlaProThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisVal 755
Db   2072 ------------------------------------------------------------ 2072
Qy    756 HisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThr 775
Db   2072 ------------------------------------------------------------ 2072
Qy    776 ThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSer 795
Db   2073 ---------------AACGCCACGGTCCGATTT------GTGTCTCTGAAGGAGCCTCT 2111
```

```
Qy        796 LeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPhe 815
              
Db       2111 ------------------------------------------------------------ 2111

Qy        816 ThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySer 835
              
Db       2111 ------------------------------------------------------------ 2111

Qy        836 ValValGluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeu 855
                    |||||:::  |||             ||||||||  |||  |||
Db       2112 ------GAGAGGACAGCCCTG---------------ACTCCCCATGTGACTGCCGGAAG 2150

Qy        856 SerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGly 875
                    :::|||
Db       2151 GAAGATGTCACC------------------------------------------------ 2162

Qy        876 GluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrp 895
                                                ||||||||
Db       2163 --------------------------------AACCATACG------------------ 2171

Qy        896 ThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAsp 915
              
Db       2171 ------------------------------------------------------------ 2171

Qy        916 ThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArg 935
              
Db       2171 ------------------------------------------------------------ 2171

Qy        936 LeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaVal 955
                                                                       :::
Db       2172 ------------------------------------------------------TCCACT 2177

Qy        956 ThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAla 975
              ||||||||::::::|||::|   :::|||:::   |||:::::::   |||:::
Db       2178 ACAGGCATTGTTATCGGCATCCACATCGGAGTCACCTGCATCATCTTCTGCGTCCTCTTT 2237

Qy        976 GlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrPro 995
                                              |||   |||          |||   |||
Db       2238 ------------------------------CTTCTGTTTGGCCA-GAGGGGCAGGGTCCTCCT 2269

Qy        996 ------------GlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerVal 1011
                          |||   ||||||||                        |||||
Db       2270 GTGTAAGGATGTGGAAAACCAGCTCTC------------------------CCCTCC--- 2302

Qy       1012 ProAlaAlaHisGluLeuGluSerLeuValHisProArgProGlnAspTrpSerProPro 1031
              |||   :::   |||   |||         :::|||   |||   :::|||:::   |||
Db       2303 CCAGGTCCCCGGAGCCAGAGGGACCCTGGTATCCTGGCCCT---GAATGGACTGAGCCG 2359

Qy       1032 ------------ProSerAspValGluAspLysAlaGluValHisSerLeuMetGlyGly 1047
                          |||::::::         ||||||              :::|||
Db       2360 AGGAGAAGGGGGCCAGCTGAGCCGAGATGAAAAGCCGGTGGACGCCAAGGAGTTGGAACA 2419

Qy       1048 SerValSerAspCysArgGlyHisSerLysArgLysIleSerTrpAlaGlnAla------ 1065
              :::||||||                                      ||||||
Db       2420 GCTGTTTCCCAC------------------------------AGCTGGGTCAGCTGCACAGCC 2452

Qy       1066 ------------------------------------------------GlyGlyProAsnTrpAla 1071
                                                            |||   |||
Db       2453 AGGCTCCACACCCACGGATCCTGCAGCCCCTGCTCCATGTGAGGAGACCCAGCTCTCCAT 2512

Qy       1072 GlySer------------------------------------------------------ 1073
              ||||||
Db       2513 GGTTCAGCTTCAAGGCTTCAACCTTGTGGCAGGGAGGACAACAGAGGCCACGTCCCCCTG 2572

Qy       1074 ---------TrpAlaGlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuThr 1090
                       ||||||              |||         |||   ||||||
Db       2573 TGCAGGGCCTGGGCC---------TGTCCCAGCACCCCAGGATATAGGCCCTGTCCCCCT 2623

Qy       1091 ArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyr 1110
              :::               ||||||   :::   ::::::      :::   :::|||
Db       2624 CAGTGAAGGACAGACCCAGCCTCCAGCAGTCGCAGCACCCCAGTGAGCTCACTGTGTTCC 2683

Qy       1111 AspGlyIleLysSerAsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGlu 1130
                 |||         :::                                 :::||||||:::
Db       2684 CGAGGGGCGGTACCCCAT-----------------------------TCTCAGGTCAAA 2713

Qy       1131 AlaGluValIle---ValHisSerAspPheGly---AlaSerLysGlyCysProAspLeu 1148
                 ||||||      |||      |||         ||||||:::|||
Db       2714 GGCGAGGTTTTTCCTGTTCTGTGGGATTTGGCGCTCGCCAGTGAAGGA------------ 2761

Qy       1149 HisLeuGlnAsp-LeuGluProGluGluProLeu------ThrAlaGluThrLeuProSe 1166
                    |||   |||:::|||:::||||||:::         |||::::::        |||
Db       2762 ------CAGGGTTCTCAAACCACAGGAACCAGTGGGTCCTACAAGCAAGGCATCTCCCTG 2815

Qy       1166 rThrSer 1168
                 :::|||
Db       2816 CTCGAGC 2822
```

RESULT 15
BB624536

```
LOCUS       BB624536                 651 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB624536 RIKEN full-length enriched, 15 days embryo male testis Mus
            musculus cDNA clone 8030498014 5', mRNA sequence.
ACCESSION   BB624536
VERSION     BB624536.1  GI:15398322
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
```

```
     source          1. .651
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="8030498O14"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="15 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 15 days embryo
                     male testis"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 185.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT       130 a    199 c    204 g    118 t
ORIGIN

Alignment Scores:
Pred. No.:              4.13e-60        Length:          651
Score:                  1032.00         Matches:         199
Percent Similarity:     98.52%          Conservative:    1
Best Local Similarity:  98.03%          Mismatches:      3
Query Match:            15.63%          Indels:          0
DB:                     10              Gaps:            0

US-09-754-997A-2 (1-1252) x BB624536 (1-651)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 100

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 160

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 220

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 280

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 340

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 400

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 460

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 520

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAAGACCAGGTG 580

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
              |||||||||||||||   |||||||||||||||||||      :::|||||||||||||||
Db        581 ACCGTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGAT 640

Qy        201 ValGlnAsp 203
              |||||||||
Db        641 GTCCAAGAC 649

Search completed: October 14, 2003, 09:25:32
Job time : 3888 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         October 14, 2003, 06:11:04 ; Search time 511 Seconds
                                             (without alignments)
                                             6613.887 Million cell updates/sec

Title:          US-09-754-997A-2
Perfect score:  6604
Sequence:       1 MARADTGRGLLVLTFCLLSA..........CPLTVSPSLPRAPVSSAQVP 1252

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09754997/runat_14102003_061059_9696/app_query.fasta_1.1415
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09754997_@CGN_1_1_0_@runat_14102003_061059_9696 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_19Jun03:*
                1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6604	100.0	6176	22	AAD10021	Mouse Nope (neighb
2	5746	87.0	3753	22	AAI67204	Nucleotide sequenc
3	5610	84.9	3741	24	ABK11101	DNA encoding human
4	4903	74.2	2796	22	AAD10022	Mouse Nope (neighb
5	4361.5	66.0	3450	24	ABZ11219	Human polynucleoti
6	3167.5	48.0	2223	22	AAI67203	Nucleotide sequenc
7	1747.5	26.5	3453	25	ABX71435	Human cDNA encodin
8	1746.5	26.4	3210	25	ABX71432	Human cDNA encodin
9	1746.5	26.4	3874	25	ABX71447	Human cDNA encodin
10	1691.5	25.6	2976	25	ABX71437	Human cDNA encodin
11	1691.5	25.6	3219	25	ABX71440	Human cDNA encodin
12	1472	22.3	825	22	AAD10023	Mouse Nope (neighb
13	1458.5	22.1	3330	23	AAS76658	DNA encoding novel
14	1384	21.0	2715	25	ABX71434	Human cDNA encodin
15	1384	21.0	2958	25	ABX71436	Human cDNA encodin
16	1329	20.1	2481	25	ABX71439	Human cDNA encodin
17	1329	20.1	2724	25	ABX71441	Human cDNA encodin
18	1144	17.3	2139	25	ABX71442	Human cDNA encodin
19	1144	17.3	2382	25	ABX71445	Human cDNA encodin
20	1067	16.2	2700	24	AAD36777	Human immunoglobul
21	958.5	14.5	4650	23	AAS70835	DNA encoding novel
22	955	14.5	5553	24	ABS76362	DNA encoding human
23	953.5	14.4	4620	25	ABT17038	Human MP21 gene ne
24	951.5	14.4	5297	24	ABV94279	Breast carcinoma r
25	951.5	14.4	5297	25	ABX10384	DNA encoding prote
26	943	14.3	4344	22	AAC90961	Human UNC-40 nucle
27	943	14.3	4608	16	AAQ80196	Deleted in colorec
28	943	14.3	4608	20	AAZ23431	Human DCC DNA. Ho
29	943	14.3	4608	25	ABT17039	Human MP21 gene DC
30	943	14.3	5186	12	AAQ12752	Deleted in Colorec
31	933.5	14.1	2695	23	AAS76339	DNA encoding novel
32	909	13.8	1875	25	ABX71443	Human cDNA encodin
33	829.5	12.6	4248	22	AAC90960	C. elegans UNC-40
34	781.5	11.8	1644	25	ABX71444	Human cDNA encodin
35	781.5	11.8	1887	25	ABX71446	Human cDNA encodin
36	768.5	11.6	6279	23	ABL14245	Drosophila melanog
37	698	10.6	2450	23	AAS76657	DNA encoding novel
38	667	10.1	7630	23	ABL30223	Drosophila melanog
39	648.5	9.8	6734	24	ABI99344	Mouse ischaemic co
40	647.5	9.8	1143	25	ABX71433	Human cDNA encodin
41	638	9.7	6000	16	AAQ86478	Human PTP-OB. Hom
42	638	9.7	6000	18	AAT85389	Human protein tyro
43	638	9.7	6000	20	AAX06095	Human protein tyro
44	638	9.7	6000	22	AAS02144	Human protein tyro
45	637.5	9.7	6374	24	AAD36781	Human immunoglobul

ALIGNMENTS

```
RESULT 1
AAD10021
ID   AAD10021 standard; cDNA; 6176 BP.
XX
AC   AAD10021;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; ss.
XX
```

```
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3759
FT                   /*tag=  a
FT                   /product= "Mouse Nope protein"
FT   sig_peptide     1..63
FT                   /*tag=  b
FT   mat_peptide     64..3756
FT                   /*tag=  c
FT                   /product= "Mature mouse Nope protein"
FT   misc_feature    1..3756
FT                   /*tag=  d
FT                   /note= "This region is specifically referred as
FT                   SEQ ID NO: 45 in the specification"
FT   misc_feature    64..2859
FT                   /*tag=  e
FT                   /note= "Extracellular domain"
FT   misc_feature    139..381
FT                   /*tag=  f
FT                   /label= Ig1
FT                   /note= "Immunoglobulin domain 1"
FT   misc_feature    463..654
FT                   /*tag=  g
FT                   /label= Ig2
FT                   /note= "Immunoglobulin domain 2"
FT   misc_feature    766..954
FT                   /*tag=  h
FT                   /label= Ig3
FT                   /note= "Immunoglobulin domain 3"
FT   misc_feature    1039..1233
FT                   /*tag=  i
FT                   /label= Ig4
FT                   /note= "Immunoglobulin domain 4"
FT   misc_feature    1285..1533
FT                   /*tag=  j
FT                   /label= Fn1
FT                   /note= "Fibronectin-type III domain 1"
FT   misc_feature    1579..1827
FT                   /*tag=  k
FT                   /label= Fn2
FT                   /note= "Fibronectin-type III domain 2"
FT   misc_feature    1888..2175
FT                   /*tag=  l
FT                   /label= Fn3
FT                   /note= "Fibronectin-type III domain 3"
FT   misc_feature    2248..2493
FT                   /*tag=  m
FT                   /label= Fn4
FT                   /note= "Fibronectin-type III domain 4"
FT   misc_feature    2542..2793
FT                   /*tag=  n
FT                   /label= Fn5
FT                   /note= "Fibronectin-type III domain 5"
FT   misc_feature    2932..3756
FT                   /*tag=  o
FT                   /note= "Intracellular domain"
FT   3'UTR           3819..5682
FT                   /*tag=  p
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAB05251.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 10; Fig 2A; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) cDNA.
XX
SQ   Sequence 6176 BP; 1316 A; 1756 C; 1634 G; 1470 T; 0 other;

Alignment Scores:
Pred. No.:              1.28e-307     Length:          6176
Score:                  6604.00       Matches:         1252
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     22            Gaps:            0

US-09-754-997A-2 (1-1252) x AAD10021 (1-6176)

Qy          1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy         21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy         41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy         61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy         81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy        121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy        141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy        161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy        181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy        201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy        221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy        241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy        261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy        281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy        301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy        321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy        341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy        361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy        381 LysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy        401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy        421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy        441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy        461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy        481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy        501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy        521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy        541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy        561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy        581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy        601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy        621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy        641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy        661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy        681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy        701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy        721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy        741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy        761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy        781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy        801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy        821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy        841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy        861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy        881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700
```

```
Qy     901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy     921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy     941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy     961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy     981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy    1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy    1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy    1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy    1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy    1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy    1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy    1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy    1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy    1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy    1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy    1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy    1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy    1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
           ||||||||||||||||||||||||||||||||||||
Db    3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 2
AAI67204
ID   AAI67204 standard; DNA; 3753 BP.
XX
AC   AAI67204;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Nucleotide sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdoch PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   P-PSDB; AAG65914.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer  -
XX
PS   Claim 2; Page 55-56; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence 3753 BP; 714 A; 1220 C; 1152 G; 667 T; 0 other;

Alignment Scores:
Pred. No.:              1.06e-266     Length:        3753
Score:                  5746.00       Matches:       1098
Percent Similarity:     92.09%        Conservative:  54
Best Local Similarity:  87.77%        Mismatches:    96
Query Match:            87.01%        Indels:        5
DB:                     22            Gaps:          2

US-09-754-997A-2 (1-1252) x AAI67204 (1-3753)

Qy         1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
```

```
Db      1 ATGGCGCGGGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 60

Qy     19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
Db     61 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 120

Qy     39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
Db    121 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 180

Qy     59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
Db    181 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 240

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db    241 GAGCACGACCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 300

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db    301 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 357

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    358 TCGTGCCTAGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 417

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    418 GCCACACTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 477

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    478 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 537

Qy    179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db    538 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 597

Qy    199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db    598 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 657

Qy    219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db    658 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACC 717

Qy    239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db    718 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 777

Qy    259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db    778 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 837

Qy    279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db    838 CAAGACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 897

Qy    299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db    898 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 957

Qy    319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGln 338
Db    958 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAG 1017

Qy    339 AlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSer 358
Db   1018 GCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCG 1077

Qy    359 GlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGly 378
Db   1078 GGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGG 1137

Qy    379 ArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAsp 398
Db   1138 CGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGAC 1197

Qy    399 AlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaPro 418
Db   1198 GCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCG 1257

Qy    419 LeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrPro 438
Db   1258 CTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCA 1317

Qy    439 LeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIle 458
Db   1318 CTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATC 1377

Qy    459 IleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPhe 478
Db   1378 ATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTT 1437

Qy    479 AlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyr 498
Db   1438 GCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTAT 1497

Qy    499 GluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeu 518
Db   1498 GAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTG 1557

Qy    519 ValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsn 538
Db   1558 GTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAAC 1617

Qy    539 ProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnVal 558
Db   1618 CCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTG 1677

Qy    559 LeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluVal 578
Db   1678 GTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTG 1737

Qy    579 ProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgVal 598
Db   1738 CGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTA 1797

Qy    599 ArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArg 618
Db   1798 CGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGG 1857

Qy    619 ThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArg 638
Db   1858 ACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAG 1917

Qy    639 AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSer 658
Db   1918 GCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCT 1977

Qy    659 GlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgPro 678
Db   1978 GGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTG 2037

Qy    679 ProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysVal 698
Db   2038 CCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTG 2097

Qy    699 LysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAla 718
Db   2098 AAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCT 2157

Qy    719 PheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaPro 738
Db   2158 TTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCG 2217
```

```
Qy     739 ThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGlu 758
Db    2218 GCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAA 2277

Qy     759 SerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLys 778
Db    2278 TCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAG 2337

Qy     779 IleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThr 798
Db    2338 ATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGCTCAGGAATGCCTCCCTGGTCACC 2397

Qy     799 TyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyr 818
Db    2398 TATTACACCAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATAC 2457

Qy     819 GluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGlu 838
Db    2458 GAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAG 2517

Qy     839 ArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeu 858
Db    2518 CGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTG 2577

Qy     859 ThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleVal 878
Db    2578 ACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTG 2637

Qy     879 GluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeu 898
Db    2638 GAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTC 2697

Qy     899 ThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyr 918
Db    2698 ACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTAC 2757

Qy     919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
Db    2758 TTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGAT 2817

Qy     939 ValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIle 958
Db    2818 GTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATC 2877

Qy     959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
Db    2878 ATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGC 2937

Qy     979 GlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnPro 998
Db    2938 CGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCC 2997

Qy     999 AlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGlu 1018
Db    2998 GCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAG 3057

Qy    1019 SerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLys 1038
Db    3058 TCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGG 3117

Qy    1039 AlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArg 1058
Db    3118 GCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGA 3177

Qy    1059 LysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGlu 1078
Db    3178 AAGATCTCCTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAG 3237

Qy    1079 LeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGly 1098
Db    3238 CTGCCCCAG---GCAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGA 3294

Qy    1099 ThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArg 1118
Db    3295 ACTGGGCAGACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGGAGG 3354

Qy    1119 LysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAsp 1138
Db    3355 AAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGAC 3414

Qy    1139 PheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluPro 1158
Db    3415 TTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCC 3474

Qy    1159 LeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAsp 1178
Db    3475 CTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCC 3534

Qy    1179 TrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThr 1198
Db    3535 TGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACC 3594

Qy    1199 CysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAla 1218
Db    3595 TGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTG 3654

Qy    1219 IleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSer 1238
Db    3655 CTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGC 3714

Qy    1239 ProSerLeuProArgAlaProValSerSerAla 1249
Db    3715 CCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3747
```

RESULT 3
ABK11101

```
ID   ABK11101 standard; DNA; 3741 BP.
XX
AC   ABK11101;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding human NOV1 protein, homologue of NOPE/PUNC Ig proteins.
XX
KW   Human; NOVX-associated disorder; developmental disorder; blood disorder;
KW   endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
KW   respiratory disorder; inflammatory disorder; reproductive disorder;
KW   neurodegenerative disorder; autoimmune disorder; infectious disease;
KW   cardiovascular disorder; cell signal processing; PUNC immunoglobulin;
KW   metabolic pathway modulation; NOPE; Ig;/NOV1; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3741
FT                   /*tag=  a
FT                   /partial
FT                   /product= "NOV1"
FT                   /note= "This sequence lacks a stop codon"
XX
PN   WO200206329-A2.
XX
PD   24-JAN-2002.
XX
PF   18-JUL-2001; 2001WO-US22709.
XX
PR   18-JUL-2000; 2000US-218870P.
PR   18-JUL-2000; 2000US-218875P.
PR   18-JUL-2000; 2000US-218901P.
PR   24-JUL-2000; 2000US-220273P.
PR   26-JUL-2000; 2000US-220912P.
PR   27-JUL-2000; 2000US-221233P.
PR   28-JUL-2000; 2000US-221650P.
XX
PA   (CURA-) CURAGEN CORP.
```

```
XX
PI   Rastelli L,  Shimkets RA,  Zerhusen B,  Malyankar UM,  Padigaru M;
XX
DR   WPI; 2002-179781/23.
DR   P-PSDB; AAU77405.
XX
PT   Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT   polypeptides, useful for treating developmental disorders, endocrine
PT   disorders, vascular disorders, infectious diseases and
PT   neurodegenerative disorders -
XX
PS   Claim 9; Page 8-9; 178pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5A,
CC   NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC   The NOVX polypeptides are related to NOPE, cadherin, interferon
CC   alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin
CC   polypeptides. The sequences of the invention are useful for identifying
CC   an agent (a cellular receptor or downstream effector) that binds to
CC   the NOVX polypeptide, or an agent that modulates it's expression or
CC   activity. They are useful for treating or preventing NOVX-associated
CC   disorders such as developmental disorders, endocrine disorders, vascular
CC   diseases, gastrointestinal disorders, respiratory disorders, inflammatory
CC   disorders, blood disorders, reproductive disorders, neurodegenerative
CC   disorders, autoimmune and immune disorders, infectious diseases,
CC   cardiovascular disorders, cancers, and other disorders related to cell
CC   signal processing and metabolic pathway modulation. The present sequence
CC   encodes the human NOV1 protein.
XX
SQ   Sequence 3741 BP; 708 A; 1209 C; 1150 G; 673 T; 1 other;

Alignment Scores:
Pred. No.:              3.53e-260      Length:         3741
Score:                  5610.00        Matches:        1086
Percent Similarity:     91.07%         Conservative:   56
Best Local Similarity:  86.60%         Mismatches:     101
Query Match:            84.95%         Indels:         13
DB:                     24             Gaps:           8

US-09-754-997A-2 (1-1252) x ABK11101 (1-3741)

Qy      1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
Db      1 ATGGCGCGGGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 60

Qy     19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
Db     61 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 120

Qy     39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
Db    121 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 180

Qy     59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
Db    181 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 240

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db    241 GAGCACGACCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 300

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db    301 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 357

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    358 TCGTGCCTAGCCCACGGNCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 417

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    418 GCCAGTCTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 477

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    478 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 537

Qy    179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db    538 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 597

Qy    199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db    598 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 657

Qy    219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db    658 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACC 717

Qy    239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db    718 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 777

Qy    259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db    778 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 837

Qy    279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db    838 ---GACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 894

Qy    299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db    895 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 954

Qy    319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgVal---LeuAlaAlaProAlaIleSer 337
Db    955 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACT 1014

Qy    338 GlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAla 357
Db   1015 CAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCG 1074

Qy    358 SerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsn 377
Db   1075 TCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAAC 1134

Qy    378 GlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGln 397
Db   1135 GGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAG 1194

Qy    398 AspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAla 417
Db   1195 GACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCG 1254

Qy    418 ProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThr 437
Db   1255 TCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACG 1314

Qy    438 ProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGln 457
Db   1315 CCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAG 1374

Qy    458 IleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGln 477
Db   1375 ATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAG 1434

Qy    478 PheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAsp 497
Db   1435 TTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGAT 1494

Qy    498 TyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAla 517
Db   1495 TATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCA 1554
```

```
Qy    518 LeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerPro 537
Db   1555 CTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCC 1614

Qy    538 AsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGln 557
Db   1615 AACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAG 1674

Qy    558 ValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGlu 577
Db   1675 GTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAG 1734

Qy    578 ValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArg 597
Db   1735 GTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGA 1794

Qy    598 ValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHis 617
Db   1795 GTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCAC 1854

Qy    618 ArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysVal 637
Db   1855 AGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTG 1914

Qy    638 ArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIle 657
Db   1915 CAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATC 1974

Qy    658 SerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArg 677
Db   1975 TCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGC 2034

Qy    678 ProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLys 697
Db   2035 CTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAA 2094

Qy    698 ValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuVal 717
Db   2095 GTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTG 2154

Qy    718 AlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAla 737
Db   2155 GCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCG 2214

Qy    738 ProThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAla 757
Db   2215 CCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCG 2274

Qy    758 GluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrVal 777
Db   2275 GAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTC 2334

Qy    778 LysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuVal 797
Db   2335 AAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTC 2394

Qy    798 ThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLys 817
Db   2395 ACCTATTAC---AGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAA 2451

Qy    818 TyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValVal 837
Db   2452 TACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTG 2511

Qy    838 GluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerPro 857
Db   2512 GAGCGCTCCACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCC 2571

Qy    858 LeuThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIle 877
Db   2572 CTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATC 2631

Qy    878 ValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeu 897
Db   2632 GTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTG 2691

Qy    898 LeuThrThr------GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAsp 915
Db   2692 CTCACCACGCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGAC 2751

Qy    916 ThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArg 935
Db   2752 ACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGC 2811

Qy    936 LeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaVal 955
Db   2812 CTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTC 2871

Qy    956 ThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAla 975
Db   2872 ACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCT 2931

Qy    976 GlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrPro 995
Db   2932 GGCCTGCGCCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCC 2991

Qy    996 GlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHis 1015
Db   2992 GGGAATCCCGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCAT 3051

Qy   1016 GluLeuGluSerLeuValHisProArgProGlnAspTrpSerProProProSerAspVal 1035
Db   3052 GAATTGGAGTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTG 3111

Qy   1036 GluAspLysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHis 1055
Db   3112 GAGGACAGGGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCAC 3171

Qy   1056 SerLysArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAla 1075
Db   3172 TCCAAAAGAAAGGTAAGT---GCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCA 3228

Qy   1076 GlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuPro 1095
Db   3229 GGCTGTGAGCTGCCCCAG---GCAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCC 3285

Qy   1096 ProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSer 1115
Db   3286 CCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGAT------CAGGGC 3339

Qy   1116 AsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleVal 1135
Db   3340 AATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTC 3399

Qy   1136 HisSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluPro 1155
Db   3400 CACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCT 3459

Qy   1156 GluGluProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGln 1175
Db   3460 GAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAG 3519

Qy   1176 GlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGlu 1195
Db   3520 GGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGAC 3579

Qy   1196 ArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnPro 1215
Db   3580 AGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCA 3639

Qy   1216 SerThrAlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeu 1235
Db   3640 GGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTA 3699

Qy   1236 ThrValSerProSerLeuProArgAlaProValSerSerAla 1249
```

```
Db       3700 GGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741

RESULT 4
AAD10022
ID   AAD10022 standard; cDNA; 2796 BP.
XX
AC   AAD10022;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) extracellular domain cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; extracellular domain; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2796
FT                   /*tag=  a
FT                   /product= "Mouse Nope extracellular domain"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAE05252.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 11; Page 77-81; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) extracellular domain cDNA.
XX
SQ   Sequence 2796 BP; 591 A; 849 C; 821 G; 535 T; 0 other;

Alignment Scores:
Pred. No.:              2.1e-226       Length:        2796
Score:                      4903.00        Matches:        932
Percent Similarity:         100.00%        Conservative:   0
Best Local Similarity:      100.00%        Mismatches:     0
Query Match:                74.24%         Indels:         0
DB:                         22             Gaps:           0

US-09-754-997A-2 (1-1252) x AAD10022 (1-2796)

Qy         22 GlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyProLeu 41
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy         42 GlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAla 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy         62 AlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHisGlu 81
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy         82 AsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAsp 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy        102 SerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCysLeu 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy        122 AlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThrLeu 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy        142 GluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArgPhe 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy        162 GluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnValThr 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy        182 ValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspVal 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy        202 GlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPhe 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy        222 SerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGln 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy        242 AspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnValVal 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy        262 MetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAspGly 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy        282 LysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAla 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy        302 GlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPhe 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy        322 AlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGlu 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy        342 AlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluPro 361
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy        362 ArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLys 381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy        382 ValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyr 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy        402 TyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaVal 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy        422 ValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSer 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy        442 SerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPhe 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy        462 SerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsn 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy        482 AsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyr 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy        502 ValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThr 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy        522 LeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAsp 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy        542 IleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyr 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy        562 LysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsn 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy        582 GluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSer 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy        602 AlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGly 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy        622 ValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMet 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy        642 GluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLys 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy        662 LeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGly 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy        682 ArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyr 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy        702 GluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLys 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy        722 HisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAsp 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy        742 LeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy        762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy        782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy        802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy        822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy        842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy        862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy        882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy        902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy        942 LeuGlnGluThrPheSerAspSerLeuAspValHis 953
              ||||||||||||||||||||||||||||||||||||
Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

```
RESULT 5
ABZ11219
ID   ABZ11219 standard; cDNA; 3450 BP.
XX
AC   ABZ11219;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 101.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
```

```
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69002.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
PS   Claim 1; SEQ ID NO 101; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3450 BP; 681 A; 1050 C; 1067 G; 652 T; 0 other;

Alignment Scores:
Pred. No.:              2.46e-200      Length:          3450
Score:                  4361.50        Matches:         822
Percent Similarity:     92.55%         Conservative:    35
Best Local Similarity:  88.77%         Mismatches:      68
Query Match:            66.04%         Indels:          1
DB:                     24             Gaps:            1

US-09-754-997A-2 (1-1252) x ABZ11219 (1-3450)

Qy        324 AlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeu 343
              :::|||:::   |||   |||||||||||||||||:::||||||||||||||||||
Db        163 TCAGCGTCGCCCCTCGCGCCCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTG 222

Qy        344 SerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgPro 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 TCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCA 282

Qy        364 AlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGln 383
              ||||||   |||||||||:::|||   ||||||||||||||||||||||||||||||||
Db        283 GCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAG 342

Qy        384 GlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGln 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAG 402

Qy        404 CysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValVal 423
              ||||||||||||||||||||||||   |||||||||||||||   |||||||||||||||
Db        403 TGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTG 462

Qy        424 ArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSer 443
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        463 CGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCT 522

Qy        444 ValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeu 463
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        523 GTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTC 582

Qy        464 HisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAsp 483
              |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        583 CACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAACAACGAC 642

Qy        484 ThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValVal 503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 ACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTACGTGGTG 702

Qy        504 AlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAsp 523
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        703 GCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACACACTGGAT 762

Qy        524 AspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArg 543
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        763 GATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGACATCAGG 822

Qy        544 ValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIle 563
              |||||||||||||||||||||   |||||||||||||||||||||:::||||||||||||
Db        823 GTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTACAAGATA 882

Qy        564 GluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThr 583
              |||||||||||||||||||||||||||||:::||||||||||||   |||||||||||||
Db        883 GAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACA 942

Qy        584 GlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGly 603
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGT 1002

Qy        604 ThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHis 623
              |||   ||||||:::|||   |||||||||||||||   ||||||||||||   :::|||
Db       1003 ACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCAC 1062

Qy        624 AsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSer 643
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db       1063 AACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATGGAGTCC 1122

Qy        644 LeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyr 663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAAACTATAT 1182

Qy        664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
              |||   |||||||||   |||||||||||||:::||||||||||||   ||||||||||||
Db       1183 TGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGGCCGTGGA 1242

Qy        684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 GACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTATGAGCTG 1302

Qy        704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db       1303 ACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAG 1362

Qy        724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
              |||||||||||||||||||||||||||||||||||||||||   ||||||:::|||
Db       1363 GATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGACATGCCT 1422

Qy        744 IleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSerSerThr 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 ATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACA 1482

Qy        764 SerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsnTyrThr 783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 TCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAACTACACT 1542

Qy        784 ValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThrSerSer 803
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1543 GTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACACCAGTTCT 1602

Qy    804 GlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGln 823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1603 GGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCAG 1662

Qy    824 SerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThrLeuPro 843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1663 TCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCCT 1722

Qy    844 AspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSerThrVal 863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 GACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTCCACGGTT 1782

Qy    864 ArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeuIleLeu 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 CGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCTGATCCTG 1842

Qy    884 TyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsn 903
          |||||:::|||||||||||||||||||||||||||||||||||||||||||:::||||||
Db   1843 TACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCAGGGAAAC 1902

Qy    904 IlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGly 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1903 ATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGG 1962

Qy    924 AlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGln 943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 GCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAG 2022

Qy    944 GluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleValGlyValCys 963
          |||         |||||||||||||||:::|||:::||||||||||||||||||||||||
Db   2023 GAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATCATCGTGGGTGTCTGC 2082

Qy    964 LeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArg 983
          ||||||||||||||||||||||||||||||||||||||||||:::|||      ||||||
Db   2083 CTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGCCGCAGCCCCCACAGG 2142

Qy    984 GluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArg 1003
          |||:::|||||||||||||||:::      ||||||||||||||||||||:::|||
Db   2143 GAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCCGCGCTGTACTCCAGA 2202

Qy   1004 AlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeuValHisPro 1023
          |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db   2203 GCTCGGCTTGGCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAGTCCCTTGTGCACCCC 2262

Qy   1024 ArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSer 1043
                |||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db   2263 CATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGGGCTGAAGTGCACAGC 2322

Qy   1044 LeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIleSerTrpAla 1063
          ||||||||||||   ||||||:::   |||   |||||||||||||||||||||||||||
Db   2323 CTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGAAAGATCTCCTGGGCT 2382

Qy   1064 GlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGlnGlySer 1083
          |||      |||   :::||||||||||||||||||||||||||||||||||      :::
Db   2383 CAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAGCTGCCCCAG---GCA 2439

Qy   1084 GlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGlyGlnThrLeu 1103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2440 GGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGAACTGGGCAGACGCTG 2499

Qy   1104 LeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLysProSerPro 1123
          ||||||||||||||||||||||||   ||||||   |||||||||||||||      |||
Db   2500 TTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGAGGAAGAAGTCACCCCCA 2559

Qy   1124 AlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGlyAlaSerLys 1143
          ||||||||||||||||||||||||||||||||||||||||||||||   |||||||
Db   2560 GCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGACTTTAGTGCATCTAAC 2619

Qy   1144 GlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThrAlaGluThr 1163
          |||   ||||||||||||||||||||||||||||||||:::||||||      |||
Db   2620 GGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCCCTGCCTCCAGAGGCT 2679

Qy   1164 LeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeuGlyArgGlu 1183
                         |||||      |||      |||||||||   ||||||   ||||||
Db   2680 CCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCCTGGCTGGACAGGGAG 2739

Qy   1184 LeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeuProGluAla 1203
          ||||||||||||:::            ||||||:::|||||||||||||||||||||||
Db   2740 TTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACCTGCTTGCCAGAGGCA 2799

Qy   1204 AlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGluGluAlaPro 1223
          ||||||||||||||||||      ||||||||||||         :::||||||   |||
Db   2800 GCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTGCTAGAGGAGACCCCT 2859

Qy   1224 GlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSerLeuProArg 1243
          |||   |||||||||   |||:::   |||||||||      ||||||   |||||||||
Db   2860 GGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGCCCAGGCTGCCCAGA 2919

Qy   1244 AlaProValSerSerAla 1249
          :::||||||||||||:::
Db   2920 TCCCCGGTCTCCTCCTCT 2937
```

RESULT 6
AAI67203
ID   AAI67203 standard; DNA; 2223 BP.
XX
AC   AAI67203;
XX
DT   11-FEB-2002 (first entry)
XX
DE   Nucleotide sequence of GSK gene Id 27142.
XX
KW   Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
KW   antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
KW   cytostatic; cerebroprotective; vasotropic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO200172961-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US09226.
XX
PR   24-MAR-2000; 2000US-192158P.
PR   28-MAR-2000; 2000US-192668P.
PR   27-APR-2000; 2000US-200166P.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PA   (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;
PI   Lai Y;
XX
DR   WPI; 2001-639223/73.
DR   P-PSDB; AAG65913.
XX
PT   Isolated polypeptides, which may be peptide hormones, which are
PT   identified by high throughput genome-based biology which identifies
PT   genes and gene products as therapeutic targets for treatment of
PT   diseases such as diabetes and cancer -
XX
PS   Claim 2; Page 54-55; 99pp; English.
XX
CC   The invention provides polypeptides (AAG65886-65918) which may be peptide
CC   hormones (including insulin, growth hormones, chemokines, cytokines,
CC   neuropeptides, integrins, kallikreins, lamins, melanins, natruiretic
CC   hormones, neuropepsin, pituitary hormones, pleiotrophins, prostaglandins,
CC   secretogranins, selectins, thromboglobulins, thymosins) identified by
CC   high throughput genome-based biology and polynucleotides (AAI67176-67208)
CC   encoding them. The polypeptides can be expressed by standard recombinant
CC   methodology. The polypeptides are useful in the treatment of disease such
CC   as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
CC   hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,

```
CC   asthma, manic depression, dementia, delirium, mental retardation,
CC   Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC   or sexual development disorders, and dysfunctions of the blood cascade
CC   system including those leading to stroke. The polynucleotides may be used
CC   as diagnostic reagents through detecting mutations in the associated gene
CC   and for chromosome localization and for tissue expression studies. The
CC   polypeptides and polynucleotides may also be used as vaccines.
XX
SQ   Sequence 2223 BP; 440 A; 660 C; 735 G; 388 T; 0 other;

Alignment Scores:
Pred. No.:              2.84e-143     Length:         2223
Score:                  3167.50       Matches:        620
Percent Similarity:     90.33%        Conservative:   34
Best Local Similarity:  85.64%        Mismatches:     54
Query Match:            47.96%        Indels:         18
DB:                     22            Gaps:           2

US-09-754-997A-2 (1-1252) x AAI67203 (1-2223)

Qy     17 LeuLeuSerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCys 36
Db     61 CTCCTCCTCTATGCAGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGT 120

Qy     37 AspGluGlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThr 56
Db    121 GGAGTGGGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGC 180

Qy     57 LeuGlyAlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThr 76
Db    181 CTGGGGGCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACC 240

Qy     77 ValLeuGluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerPro 96
Db    241 CTGCTGGAGCACGACCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCA 300

Qy     97 LeuGluGlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGly 116
Db    301 CTAGCACCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGC 357

Qy    117 SerTyrSerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValVal 136
Db    358 AACTATTCGTGCCTAGCCCACGGCCCCCCTGGAGTGCTGGCCAGCCAGACTGCTGTCGTC 417

Qy    137 LysLeuAlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsn 156
Db    418 AAGCTTGCCACACTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAAC 477

Qy    157 GlyThrAlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGlu 176
Db    478 GGGACAGCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAG 537

Qy    177 LysAspGlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeu 196
Db    538 AAGGACCAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTT 597

Qy    197 GlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSer 216
Db    598 CAGATCCTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCA 657

Qy    217 AlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGlu 236
Db    658 GCTCGCCAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCG 717

Qy    237 AlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSer 256
Db    718 TCCACCAGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCT 777

Qy    257 GlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrp 276
Db    778 GGCCAGAGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGG 837

Qy    277 ValArgGlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeu 296
Db    838 GTCCGACAAGACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTA 897

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    898 CTAATTGCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCC 957

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    958 CGCACGCGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATC 1017

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db   1018 ACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGC 1077

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db   1078 GCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCC 1137

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db   1138 AACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTG 1197

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db   1198 CAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCC 1257

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db   1258 GCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCT 1317

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1318 ACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAG 1377

Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1378 CAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATAC 1437

Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1438 CAGTTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACA 1497

Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1498 GATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCA 1557

Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1558 GCACTGGTGCACACACTGGATGAT------GGTAGGGCCTCTGAACTCGCAGTGGGCAGC 1611

Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1612 ------------------------------------------TTGGGCCTGAGCAATGGG 1629

Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1630 CAGGTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACT 1689

Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1690 GAGGTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTAT 1749

Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1750 CGAGTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCAT 1809

Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1810 CACAGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAG 1869

Qy    637 ValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGln 656
```

```
Db       1870 GTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAG 1929

Qy        657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
              ||||||||||||||||||||||||||  |||||||||   |||||||||||||:::||||||
Db       1930 ATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGAT 1989

Qy        677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
              |||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1990 CGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAG 2049

Qy        697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db       2050 AAAGTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTC 2109

Qy        717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2110 GTGGCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAG 2169

Qy        737 AlaProThrPro 740
              ||||||   |||
Db       2170 GCGCCGGCACCA 2181
```

RESULT 7
ABX71435

```
ID   ABX71435 standard; cDNA; 3453 BP.
XX
AC   ABX71435;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP4.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54203.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Claim 1; Column 31-36; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3453 BP; 979 A; 821 C; 769 G; 883 T; 1 other;

Alignment Scores:
Pred. No.:              5.79e-75     Length:        3453
Score:                  1747.50      Matches:       397
Percent Similarity:     54.95%       Conservative:  169
Best Local Similarity:  38.54%       Mismatches:    381
Query Match:            26.46%       Indels:        83
DB:                     25           Gaps:          18

US-09-754-997A-2 (1-1252) x ABX71435 (1-3453)

Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
              |||:::|||:::                  |||||                |||||||||
Db         43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                             :::||||||     :::   |||       |||   :::
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
              :::         ||||||||||||||                            |||   :::
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
              ||||||||   |||:::|||      :::   |||::::::   :::   :::|||
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
              |||||||||||:::::::|||           |||   |||            |||::::::
Db        268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                   ||||||   |||   ||||||||    :::       |||
Db        319 -----------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::   ||||||   |||   :::   |||:::|||:::      |||      :::
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
              |||   |||         |||   |||   |||   |||||||||   |||
Db        421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
                 |||      :::||||||||||||   ::::::   |||:::|||
Db        481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
              |||:::      ||||||      :::|||||||||   ||||||         |||:::
Db        541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
              |||:::||||||||||||:::|||         |||   :::|||   |||   |||||||||
Db        601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
              |||||||||            :::   |||:::                     |||:::
Db        661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCAACAATTATA 714

Qy        247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
              |||   |||:::|||   |||                  |||   ||||||:::||||||:::|||
Db        715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy        267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
```

```
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834
Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954
Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014
Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074
Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128
Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188
Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248
Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308
Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368
Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548
Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668
Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728
Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788
Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845
Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902
Qy    645 ValValSerTrpGlnProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962
Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   1963 TACAAGGAAGAAGGGCAGCAGGAGAAT------------------------------ 1989
Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   1990 -----------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031
Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091
Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2092 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136
Qy    744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196
Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256
Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db   2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316
Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376
Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436
Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496
Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556
Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616
Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676
Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736
Qy    940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db   2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796
Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db   2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856
Qy    960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db   2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916
```

```
Qy        980 SerSerHisArgGluAla------------------------------------LeuPro 987
                  |||     |||::::::                                 ||||||
Db       2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976

Qy        988 GlyLeuSerSerSerGlyThrProGlyAsn 997
                       |||:::|||         ||||||
Db       2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006

RESULT 8
ABX71432
ID   ABX71432 standard; cDNA; 3210 BP.
XX
AC   ABX71432;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP1.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein;
KW   SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       76
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       706
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54200.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 13-18; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3210 BP; 895 A; 761 C; 716 G; 836 T; 2 other;

Alignment Scores:
Pred. No.:               5.98e-75      Length:         3210
Score:                   1746.50       Matches:        397
Percent Similarity:      54.95%        Conservative:   169
Best Local Similarity:   38.54%        Mismatches:     381
Query Match:             26.45%        Indels:         83
DB:                      25            Gaps:           18

US-09-754-997A-2 (1-1252) x ABX71432 (1-3210)

Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
                |||:::|||:::                  ||||||                  |||||||||
Db         43 GGGATGCTGCTCCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                             :::||||||       :::     |||       |||    :::
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
                :::      ||||||||||||||                             |||   :::
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
              |||||||||      |||:::|||          :::      |||::::::    :::      :::|||
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
              ||||||||||||::::::|||               |||   |||         |||::::::
Db        268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                    ||||||   |||   |||||||||      :::      |||
Db        319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::      ||||||   |||   :::    |||:::|||:::             |||    :::
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
              |||     |||          |||   |||   |||   |||||||||   |||
Db        421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
                 |||          :::|||||||||||      ::::::    |||:::|||
Db        481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
              |||:::         ||||||     :::|||||||||   ||||||              |||:::
Db        541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
              |||:::|||||||||:::|||            |||     :::|||     |||   |||||||||
Db        601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
              |||||||||                  :::      |||:::                  |||:::
Db        661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy        247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
              |||      |||:::|||   |||               |||   ||||||:::||||||:::|||
Db        715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy        267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
              :::      :::|||   |||   :::||||||   |||   |||   |||   |||
Db        775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy        287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
```

```
Db   835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
Qy   306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db   895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954
Qy   326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db   955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014
Qy   346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db  1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074
Qy   366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db  1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128
Qy   386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db  1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188
Qy   406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db  1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248
Qy   426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db  1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308
Qy   446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db  1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368
Qy   466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db  1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy   486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db  1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy   506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db  1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548
Qy   526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db  1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608
Qy   546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db  1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668
Qy   566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db  1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728
Qy   586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db  1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788
Qy   606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db  1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845
Qy   626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db  1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902
Qy   645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663

Db  1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962
Qy   664 TrpGlyGluValGlyThrGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db  1963 TACAAGGAAGAAGGGCAGCAGGAGAAT------------------------------- 1989
Qy   684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db  1990 -----------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031
Qy   704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db  2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091
Qy   724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db  2092 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136
Qy   744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db  2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196
Qy   760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db  2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256
Qy   780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db  2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316
Qy   800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db  2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376
Qy   820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db  2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436
Qy   840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db  2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496
Qy   860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db  2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556
Qy   880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db  2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616
Qy   900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db  2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676
Qy   920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db  2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736
Qy   940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db  2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796
Qy   950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db  2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856
Qy   960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db  2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916
Qy   980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db  2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976
```

```
Qy        988 GlyLeuSerSerSerGlyThrProGlyAsn 997
Db       2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006

RESULT 9
ABX71447
ID   ABX71447 standard; cDNA; 3874 BP.
XX
AC   ABX71447;
XX
DT   06-MAR-2003  (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 107-112; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3874 BP; 1077 A; 938 C; 897 G; 960 T; 2 other;

Alignment Scores:
Pred. No.:              7.31e-75      Length:         3874
Score:                  1746.50       Matches:        397
Percent Similarity:     54.95%        Conservative:   169
Best Local Similarity:  38.54%        Mismatches:     381
Query Match:            26.45%        Indels:         83
DB:                     25            Gaps:           18

US-09-754-997A-2 (1-1252) x ABX71447 (1-3874)

Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db        233 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 286

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db        287 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 346

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db        347 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 397

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db        398 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 457

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db        458 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 508

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db        509 -----------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 550

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db        551 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 610

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db        611 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 670

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db        671 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 730

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db        731 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 790

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db        791 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 850

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db        851 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 904

Qy        247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db        905 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 964

Qy        267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db        965 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 1024

Qy        287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db       1025 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 1084

Qy        306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db       1085 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 1144

Qy        326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db       1145 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1204

Qy        346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db       1205 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1264
```

```
Qy   366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db  1265 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1318

Qy   386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db  1319 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1378

Qy   406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db  1379 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1438

Qy   426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db  1439 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1498

Qy   446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db  1499 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1558

Qy   466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db  1559 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1618

Qy   486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db  1619 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1678

Qy   506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db  1679 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1738

Qy   526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db  1739 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1798

Qy   546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db  1799 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1858

Qy   566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db  1859 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1918

Qy   586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db  1919 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1978

Qy   606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db  1979 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 2035

Qy   626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db  2036 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 2092

Qy   645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db  2093 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 2152

Qy   664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db  2153 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 2179

Qy   684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db  2180 ------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2221

Qy   704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db  2222 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2281

Qy   724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db  2282 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2326

Qy   744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db  2327 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2386

Qy   760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db  2387 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2446

Qy   780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db  2447 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2506

Qy   800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db  2507 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2566

Qy   820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db  2567 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2626

Qy   840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db  2627 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2686

Qy   860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db  2687 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2746

Qy   880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db  2747 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2806

Qy   900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db  2807 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2866

Qy   920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db  2867 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2926

Qy   940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db  2927 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2986

Qy   950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db  2987 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 3046

Qy   960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db  3047 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 3106

Qy   980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db  3107 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 3166

Qy   988 GlyLeuSerSerSerGlyThrProGlyAsn 997
Db  3167 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3196
```

RESULT 10
ABX71437
ID   ABX71437 standard; cDNA; 2976 BP.
XX
AC   ABX71437;
XX
DT   06-MAR-2003 (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP6.

```
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW,  Scoville J,  Turner CA,  Friedrich G,  Abuin A;
PI   Zambrowicz B,  Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54205.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays  -
XX
PS   Disclosure; Column 49-52; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 2976 BP; 844 A; 697 C; 653 G; 781 T; 1 other;

Alignment Scores:
Pred. No.:              2.4e-72      Length:         2976
Score:                  1691.50      Matches:        375
Percent Similarity:     55.79%       Conservative:   160
Best Local Similarity:  39.10%       Mismatches:     347
Query Match:            25.61%       Indels:         77
DB:                     25           Gaps:           15

US-09-754-997A-2 (1-1252) x ABX71437 (1-2976)

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107

Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167

Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227
```

```
Qy        497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
              :::|||   ||||||:::||||||||       :::|||||||||:::    |||
Db       1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287

Qy        517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
                 :::||||||:::||||||||             |||:::::::::|||:::|||
Db       1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347

Qy        537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
                :::|||:::||||||  :::::|||||||||:::|||:::                |||
Db       1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGC 1407

Qy        557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
              ||||||:::   |||::::::   :::   |||    |||:::    :::
Db       1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467

Qy        577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
              |||:::||||||         :::      |||      |||:::|||:::   ||||||
Db       1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527

Qy        597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
                |||||||||:::|||   |||      |||   |||      |||   |||
Db       1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587

Qy        617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
              ||||||||||||         |||   :::|||   ||||||
Db       1588 CATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641

Qy        637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
              :::                  ::::::   |||   ||||||                |||
Db       1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701

Qy        656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAsp 674
                    |||   ||||||||||||||:::   |||   |||   :::|||
Db       1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755

Qy        675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694
                                                         ||||||:::   |||
Db       1756 ---------------------------------------------GGGCCCATTTTCTTG 1770

Qy        695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714
                     |||         |||   |||:::   |||   |||   |||   |||   |||
Db       1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830

Qy        715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734
              :::|||:::|||:::|||   :::||||||||||   |||
Db       1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875

Qy        735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750
              :::            ||||||      |||                     ||||||
Db       1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935

Qy        751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770
              ||||||   |||:::::::|||::::::|||:::|||:::||||||:::|||   |||:::
Db       1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995

Qy        771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790
              :::|||   ||||||      :::|||:::|||||||||:::|||      |||   |||
Db       1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055

Qy        791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810
              |||:::||||||||||||||||||   |||               :::|||:::
Db       2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAA 2115

Qy        811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830
              ||||||:::|||      |||||||||||||||||||:::   |||         :::
Db       2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175

Qy        831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
                 |||:::         ||||||   ||||||||||||:::   |||:::   ||||||
Db       2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235

Qy        851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
                 ::::::::::::   :::          |||       :::   |||   ||||||
Db       2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295

Qy        871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
                 |||         :::      |||   |||||||||::::::
Db       2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355

Qy        891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
                    :::|||   :::|||      ||||||   |||      |||   :::
Db       2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415

Qy        911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
                 |||   :::          |||      |||   |||      |||||||||
Db       2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475

Qy        931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949
              ||||||||||||         ::::::      |||   :::      |||:::|||
Db       2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535

Qy        950 ---------------------------------------------------------Leu 950
                                                                       |||
Db       2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595

Qy        951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
                 |||      :::::::||||||||      |||||||||   :::   |||   |||:::|||
Db       2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655

Qy        971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985
                 |||:::         :::   :::|||      |||::::::
Db       2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715

Qy        986 ---------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
                                   ||||||   |||:::|||         ||||||
Db       2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 11
ABX71440
ID   ABX71440 standard; cDNA; 3219 BP.
XX
AC   ABX71440;
XX
DT   06-MAR-2003 (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP9.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW, Scoville J, Turner CA, Friedrich G, Abuin A;
PI   Zambrowicz B, Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54208.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays -
XX
PS   Disclosure; Column 65-68; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a

```
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 3219 BP; 927 A; 757 C; 706 G; 828 T; 1 other;

Alignment Scores:
Pred. No.:              2.61e-72      Length:         3219
Score:                  1691.50       Matches:        375
Percent Similarity:     55.79%        Conservative:   160
Best Local Similarity:  39.10%        Mismatches:     347
Query Match:            25.61%        Indels:         77
DB:                     25            Gaps:           15

US-09-754-997A-2 (1-1252) x ABX71440 (1-3219)

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC-----------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107

Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167

Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227

Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287

Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347

Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGC 1407

Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467

Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527

Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587

Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1588 CATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641

Qy    637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
```

```
Db     1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701

Qy      656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAsp 674

Db     1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755

Qy      675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694

Db     1756 ------------------------------------------------GGGCCCATTTTCTTG 1770

Qy      695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714

Db     1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830

Qy      715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734

Db     1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875

Qy      735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750

Db     1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935

Qy      751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770

Db     1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995

Qy      771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790

Db     1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055

Qy      791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810

Db     2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAA 2115

Qy      811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830

Db     2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175

Qy      831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850

Db     2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235

Qy      851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870

Db     2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295

Qy      871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890

Db     2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355

Qy      891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910

Db     2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415

Qy      911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930

Db     2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475

Qy      931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949

Db     2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535

Qy      950 ---------------------------------------------------------Leu 950

Db     2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595

Qy      951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970

Db     2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655

Qy      971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985

Db     2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715

Qy      986 ---------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997

Db     2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 12
AAD10023
ID   AAD10023 standard; cDNA; 825 BP.
XX
AC   AAD10023;
XX
DT   12-SEP-2001 (first entry)
XX
DE   Mouse Nope (neighbour of punc ell) intracellular domain cDNA.
XX
KW   Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW   gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW   tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW   obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW   polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW   renal anomaly; cardiovascular anomaly; intracellular domain; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..825
FT   /*tag= a
FT   /product= "Mouse Nope intracellular domain"
FT   /note= "CDS does not include start and stop codon"
FT   /partial
XX
PN   WO200149714-A2.
XX
PD   12-JUL-2001.
XX
PF   26-OCT-2000; 2000WO-US29698.
XX
PR   04-JAN-2000; 2000US-0174496.
PR   19-MAY-2000; 2000US-0205789.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Salbaum JM;
XX
DR   WPI; 2001-441846/47.
DR   P-PSDB; AAE05253.
XX
PT   Murine Nope polypeptides and nucleic acids useful for preventing,
PT   diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
XX
PS   Claim 11; Page 83-85; 99pp; English.
XX
CC   The present invention relates to Nope (neighbour of punc ell) which is
CC   used in the prevention, treatment and diagnosis of diseases associated
CC   with inappropriate Nope expression such as cancers especially colonic
CC   cancer and genetic disorders, as Nope is thought to be a tumour
CC   suppressor. Nope gene is located on chromosome 9 and is used in gene
CC   therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC   diseases by rectifying mutations or deletions in a patient's genome that
CC   affect the activity of Nope by expressing inactive proteins or to
CC   supplement the patients own production of Nope polypeptides. Nope gene
CC   is used to study the expression and function of Nope polypeptides and
CC   their role in metabolism through the creation of transgenic animal
CC   models. The anti-Nope antibodies and Nope antagonists may also be used
CC   to down regulate Nope expression and activity for the treatment of
CC   Bardet-Biedl syndrome which is an autosomal recessive disorder
CC   characterised by mental retardation, obesity, polydactyly, retinitis
CC   pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC   high incidence of hypertension, diabetes mellitus and renal and
CC   cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC   of punc ell) intracellular domain cDNA.
XX

```
SQ   Sequence 825 BP; 174 A; 271 C; 236 G; 144 T; 0 other;

Alignment Scores:
Pred. No.:               2.06e-62      Length:         825
Score:                   1472.00       Matches:        275
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             22.29%        Indels:         0
DB:                      22            Gaps:           0

US-09-754-997A-2 (1-1252) x AAD10023 (1-825)

Qy        978 ArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 60

Qy        998 ProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeu 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 120

Qy       1018 GluSerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAsp 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 180

Qy       1038 LysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLys 1057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 240

Qy       1058 ArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCys 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 300

Qy       1078 GluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAla 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 360

Qy       1098 GlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGly 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 420

Qy       1118 ArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSer 1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 480

Qy       1138 AspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGlu 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 540

Qy       1158 ProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAla 1177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 600

Qy       1178 AspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeu 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 660

Qy       1198 ThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThr 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 720

Qy       1218 AlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrVal 1237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 780

Qy       1238 SerProSerLeuProArgAlaProValSerSerAlaGlnValPro 1252
              |||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 825

RESULT 13
AAS76658.
ID   AAS76658 standard; cDNA; 3330 BP.
XX
AC   AAS76658;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12462.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12471.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 12462; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3330 BP; 971 A; 775 C; 746 G; 838 T; 0 other;

Alignment Scores:
Pred. No.:               4.05e-61      Length:         3330
Score:                   1458.50       Matches:        341
Percent Similarity:      50.58%        Conservative:   139
Best Local Similarity:   35.93%        Mismatches:     324
Query Match:             22.09%        Indels:         145
DB:                      23            Gaps:           16

US-09-754-997A-2 (1-1252) x AAS76658 (1-3330)

Qy        131 SerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHisProGluSer 150
              |||            :::|||||         ::::::         |||   :::   |||   |||
Db        226 AGTAGCTCAATGTTAGTGAAACACCATTCTATTTCTGCATTTGAAGTCCAGCCAATTTCC 285

Qy        151 GlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGlyLeuProAla 170
                    |||   |||   |||   |||||||||   |||               |||
```

```
Db    286 ACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCCCACCCTCCT 345
Qy    171 ProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluProArgLeuIle 189
Db    346 GCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGACAGGATAACT 405
Qy    190 ThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAlaGlySerTyr 209
Db    406 GCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCTGGAAATTAT 465
Qy    210 ArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSerLeuThrVal 229
Db    466 CGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTG 525
Qy    230 AlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleValAlaAlaPro 249
Db    526 ATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCAACAATTATAGCAGGTCCA 579
Qy    250 GluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAlaSerAlaAsp 269
Db    580 CAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCCACAGGAAAT 639
Qy    270 ProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr---AspVal 288
Db    640 CCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTCTTTAATACT 699
Qy    289 IleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyVal 308
Db    700 CGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACATGCTGGAGTA 759
Qy    309 TyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeu 328
Db    760 TATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATGGCAACTTTA 819
Qy    329 ArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAla 348
Db    820 ACTGTATTA---------------------------GAAAGTTTAACAAGGCCTCGAGCT 852
Qy    349 SerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeu 368
Db    853 GGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTG 912
Qy    369 HisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGlySer 388
Db    913 AAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------AACAGTAAA 966
Qy    389 LeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsn 408
Db    967 TTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATGGCTGAGAAT 1026
Qy    409 SerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGluGlyLeuPro 428
Db   1027 AGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAAGACAGACCC 1086
Qy    429 SerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAlaTrp 448
Db   1087 AGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGG 1146
Qy    449 GluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLysAla 468
Db   1147 GAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTACATGAAAGCA 1206
Qy    469 ArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGln 488
Db   1207 GAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACTCATTATATT 1266
Qy    489 ValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeu 508
Db   1267 ATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATATATGCCAATG 1326
Qy    509 GlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspValProSerAla 528
Db   1327 GGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGAT------------ 1374

Qy    529 AlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeuPro 548
Db   1374 ------------------------------------------------------------ 1374
Qy    549 LeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeuGly 568
Db   1375 ------GCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGT 1428
Qy    569 LysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeuAsn 588
Db   1429 ACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAA 1488
Qy    589 SerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyr 608
Db   1489 GGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTG 1548
Qy    609 GlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSerHisVal 628
Db   1549 GGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCC 1605
Qy    629 ProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeuValValSer 647
Db   1606 CCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGG 1662
Qy    648 TrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyrTrpGlyGlu 666
Db   1663 TGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAA 1722
Qy    667 ValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGlnAla 686
Db   1723 GAAGGGCAGCAGGAGAAT------------------------------------------ 1740
Qy    687 TrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGlnLeu 706
Db   1741 ---------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTCAGTGGCTTA 1791
Qy    707 Val--------------------------------------------------------- 707
Db   1792 GCGTTTTTCCTGCTCCTCTGCTGCTGGGCCAGACTTGAGCTGGCCCCAGATTCTATGTCA 1851
Qy    707 ------------------------------------------------------------ 707
Db   1852 GAGGCCATGACCCAGGTGTCAGAGCAGGCCGTACCGCTGGCCTCTTCTCTGAATAATTAT 1911
Qy    708 ---------ProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAsp 724
Db   1912 ATGGGGGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGAT 1971
Qy    725 GlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIle 744
Db   1972 GGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCT 2016
Qy    745 GlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsn 760
Db   2017 GTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAAC 2076
Qy    761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
Db   2077 ACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATT 2136
Qy    781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
Db   2137 AACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTT 2196
Qy    801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
Db   2197 CAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTT 2256
Qy    821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
Db   2257 GCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCT 2316
```

```
Qy       841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
             |||||||||:::   |||:::   ||||||       ::::::::::::   :::
Db      2317 ACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGAT 2376

Qy       861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
                |||       :::  |||  ||||||      |||          :::      |||
Db      2377 GACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTAT 2436

Qy       881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
                ||||||||::::::                          :::|||   :::|||
Db      2437 ACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGT 2496

Qy       901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
              |||||   |||     |||   :::         |||    :::          |||
Db      2497 GAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTC 2556

Qy       921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
             |||:::   |||      |||||||||   ||||||||||            ::::::
Db      2557 AAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCA 2616

Qy       941 ThrLeuGlnGluThrPheSerAspSer--------------------------------- 949
                |||   :::      |||:::|||
Db      2617 GTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGAT 2676

Qy       950 -------------------------LeuAspValHisAlaValThrGlyIleIleVal 960
                                       ||||||      ::::::|||||||||   |||
Db      2677 GCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTA 2736

Qy       961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
             ||||||    :::   |||   |||:::|||   |||:::          :::   :::|||
Db      2737 GGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGT 2796

Qy       981 SerHisArgGluAla---------------------------------LeuProGly 988
                   |||::::::                                 ||||||
Db      2797 AAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGT 2856

Qy       989 LeuSerSerSerGlyThrProGlyAsn 997
                |||:::|||       ||||||
Db      2857 ACCAGTGCCTCCTTAGCTAGTGGAAAT 2883
```

RESULT 14
ABX71434
ID   ABX71434 standard; cDNA; 2715 BP.
XX
AC   ABX71434;
XX
DT   06-MAR-2003 (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP3.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW, Scoville J, Turner CA, Friedrich G, Abuin A;
PI   Zambrowicz B, Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54202.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays -
XX
PS   Disclosure; Column 25-28; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 2715 BP; 772 A; 634 C; 590 G; 717 T; 2 other;

Alignment Scores:

Pred. No.:	1.22e-57	Length:	2715
Score:	1384.00	Matches:	334
Percent Similarity:	46.11%	Conservative:	140
Best Local Similarity:	32.49%	Mismatches:	310
Query Match:	20.96%	Indels:	244
DB:	25	Gaps:	14

US-09-754-997A-2 (1-1252) x ABX71434 (1-2715)

```
Qy         9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
              |||:::|||:::                 ||||||               |||||||||
Db        43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy        29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                            :::||||||       :::   |||      |||   :::
Db        97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy        48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
              :::         |||||||||||||||                        |||   :::
Db       157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy        68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
              |||||||||   |||:::|||       :::   |||::::::   :::   :::|||
Db       208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy        88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
              ||||||||||||::::::|||         |||   |||         |||::::::
Db       268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy       108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                  ||||||   |||   |||||||||   :::      |||
Db       319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy       128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::   ||||||   |||   :::   |||:::|||:::         |||   :::
Db       361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy       148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
              |||   |||      |||   |||   |||   |||||||||   |||
Db       421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy       168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
              |||         :::||||||||||||   ::::::   |||:::|||
Db       481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540
```

```
Qy    187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db    541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy    207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db    601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy    227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
Db   1554 ------------------------------------------------------------ 1554

Qy    646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   1554 ------------------------------------------------------------ 1554

Qy    666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   1554 ------------------------------------------------------------ 1554

Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   1554 ------------------------------------------------------------ 1554

Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602

Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   1603 TATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647

Qy    746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db   1648 CGTGATCGCATGGTCCCTCCTCCACCACCACCCACCATCTCTATGCGAAGGCTAACACC 1707

Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db   1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767

Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db   1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827

Qy    802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db   1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887

Qy    822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db   1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947

Qy    842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db   1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007

Qy    862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
Db   2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067

Qy    882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db   2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127

Qy    902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
```

```
                 |||  |||      |||   :::     |||   :::        |||     |||
Db       2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187

Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
              :::  |||      ||||||||  |||||||||||          ::::::
Db       2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247

Qy        942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
              |||   :::     |||:::|||
Db       2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307

Qy        950 ------------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
                                      ||||||     ::::::|||||||||  ||||||
Db       2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367

Qy        962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
              |||    :::    |||    |||:::|||  |||:::        :::    :::|||
Db       2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427

Qy        982 HisArgGluAla-----------------------------------LeuProGlyLeu 989
                 |||::::::                                   ||||||
Db       2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487

Qy        990 SerSerSerGlyThrProGlyAsn 997
              |||:::|||         ||||||
Db       2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

RESULT 15
ABX71436
ID   ABX71436 standard; cDNA; 2958 BP.
XX
AC   ABX71436;
XX
DT   06-MAR-2003 (first entry)
XX
DE   Human cDNA encoding a novel tyrosine phosphatase, NHP5.
XX
KW   Human; ss; gene; chromosome 15; chromosome 3; tyrosine phosphatase;
KW   Ig super family protein; gene therapy; NHP; novel human protein.
XX
OS   Homo sapiens.
XX
PN   US6465632-B1.
XX
PD   15-OCT-2002.
XX
PF   08-JUN-2001; 2001US-0877730.
XX
PR   09-JUN-2000; 2000US-210607P.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW, Scoville J, Turner CA, Friedrich G, Abuin A;
PI   Zambrowicz B, Sands AT;
XX
DR   WPI; 2003-147071/14.
DR   P-PSDB; ABU54204.
XX
PT   Novel isolated nucleic acid which encodes a novel human protein that
PT   shares sequence similarity with animal phosphatases, that is useful for
PT   generating antibodies, and as reagents in diagnostic assays -
XX
PS   Disclosure; Column 41-44; 58pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule comprising a
CC   nucleotide sequence encoding a novel human protein (NHP) that shares
CC   sequence similarity with animal phosphatases (in particular tyrosine
CC   phosphatases and is a member if the Ig superfamily). The disclosed NHP
CC   polynucleotide sequences (from genes located on either chromosome 15
CC   or chromosome 3) are useful for identifying coding sequences and in the
CC   identification of biologically relevant splice junctions. The NHP
CC   polynucleotide sequences are useful in gene therapy, for detecting
CC   mutant NHPs or inappropriately expressed NHPs, for the diagnosis of
CC   disease, for screening drugs effective in treatment of symptomatic or
CC   phenotypic manifestations of perturbing the normal function of NH.
CC   Complementary sequences of the NHP polynucleotides used in conjunction
CC   with PCR to screen libraries, isolate clones and prepare cloning and
CC   sequencing templates. Labeled NHP nucleotide probes can be used to screen
CC   a human genomic library which is helpful for identifying polymorphisms,
CC   determining the genomic structure of a given locus/allele and designing
CC   diagnostic tests. The probe sequences also have use in defining and
CC   monitoring both drug action and toxicity. Oligonucleotides
CC   complementary to NHPs may encode or act as NHP antisense molecules, or
CC   may be used as part of ribozyme and/or triple helix sequences.
CC   The present sequence encodes an NHP.
XX
SQ   Sequence 2958 BP; 855 A; 694 C; 643 G; 764 T; 2 other;

Alignment Scores:

Pred. No.:	1.33e-57	Length:	2958
Score:	1384.00	Matches:	334
Percent Similarity:	46.11%	Conservative:	140
Best Local Similarity:	32.49%	Mismatches:	310
Query Match:	20.96%	Indels:	244
DB:	25	Gaps:	14

US-09-754-997A-2 (1-1252) x ABX71436 (1-2958)

```
Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
              |||:::|||:::                                   |||||||||
Db         43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                          :::||||||      :::   |||      |||   :::
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
              :::      ||||||||||||||                           |||   :::
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
              ||||||||   |||:::|||      :::   |||::::::   :::   :::|||
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
              ||||||||||||::::::|||         |||   |||            |||::::::
Db        268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                ||||||   |||   ||||||||      :::      |||
Db        319 ------------------GATGAAGGATTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::   ||||||   |||   :::   |||:::|||:::          |||   :::
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
              |||   |||         |||   |||   |||   |||||||||   |||
Db        421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
                 |||         :::||||||||||||      ::::::    |||:::|||
Db        481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
              |||:::         ||||||      :::|||||||||   ||||||      |||:::
Db        541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
              |||:::||||||||||:::|||            |||   :::|||   |||   |||||||||
Db        601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
              |||||||||               :::   |||:::                  |||:::
Db        661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714
```

```
Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774
Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834
Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954
Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014
Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074
Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128
Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188
Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248
Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308
Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368
Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548
Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554
Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554
Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554
Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554
Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
Db   1554 ------------------------------------------------------------ 1554
Qy    646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   1554 ------------------------------------------------------------ 1554
Qy    666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   1554 ------------------------------------------------------------ 1554
Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   1554 ------------------------------------------------------------ 1554
Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602
Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   1603 TATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647
Qy    746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db   1648 CGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACC 1707
Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db   1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767
Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db   1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827
Qy    802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db   1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887
Qy    822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db   1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947
Qy    842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db   1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007
Qy    862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
Db   2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067
Qy    882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db   2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127
Qy    902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
Db   2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187
Qy    922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
Db   2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247
Qy    942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
Db   2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307
Qy    950 ------------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
```

```
Db     2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367

Qy      962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
            |||   :::   |||   |||:::|||   |||:::         :::   :::|||
Db     2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427

Qy      982 HisArgGluAla------------------------------------LeuProGlyLeu 989
               |||::::::                                        ||||||
Db     2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487

Qy      990 SerSerSerGlyThrProGlyAsn 997
            |||:::|||         ||||||
Db     2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

Search completed: October 14, 2003, 06:22:47
Job time : 680 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         October 14, 2003, 06:11:10 ; Search time 872 Seconds
                                            (without alignments)
                                            3725.578 Million cell updates/sec

Title:          US-09-754-997A-2
Perfect score:  6604
Sequence:       1 MARADTGRGLLVLTFCLLSA..........CPLTVSPSLPRAPVSSAQVP 1252

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters:      3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09754997/runat_14102003_061101_9762/app_query.fasta_1.1415
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09754997_@CGN_1_1_243_@runat_14102003_061101_9762
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                16:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                17:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6604	100.0	3756	10	US-09-754-997A-45	Sequence 45, Appl
2	6604	100.0	6176	10	US-09-754-997A-1	Sequence 1, Appli
3	5746	87.0	3753	12	US-10-239-663-29	Sequence 29, Appl
4	5610	84.9	3741	10	US-09-908-193-1	Sequence 1, Appli
5	4903	74.2	2796	10	US-09-754-997A-3	Sequence 3, Appli
6	3167.5	48.0	2223	12	US-10-239-663-28	Sequence 28, Appl
7	1747.5	26.5	3453	14	US-10-231-353-7	Sequence 7, Appli
8	1746.5	26.4	3210	14	US-10-231-353-1	Sequence 1, Appli
9	1746.5	26.4	3874	14	US-10-231-353-31	Sequence 31, Appl
10	1691.5	25.6	2976	14	US-10-231-353-11	Sequence 11, Appl
11	1691.5	25.6	3219	14	US-10-231-353-17	Sequence 17, Appl
12	1472	22.3	825	10	US-09-754-997A-5	Sequence 5, Appli
13	1384	21.0	2715	14	US-10-231-353-5	Sequence 5, Appli
14	1384	21.0	2958	14	US-10-231-353-9	Sequence 9, Appli
15	1329	20.1	2481	14	US-10-231-353-15	Sequence 15, Appl
16	1329	20.1	2724	14	US-10-231-353-19	Sequence 19, Appl
17	1144	17.3	2139	14	US-10-231-353-21	Sequence 21, Appl
18	1144	17.3	2382	14	US-10-231-353-27	Sequence 27, Appl
19	951.5	14.4	5297	12	US-10-007-926A-270	Sequence 270, App
20	909	13.8	1875	14	US-10-231-353-23	Sequence 23, Appl
21	781.5	11.8	1644	14	US-10-231-353-25	Sequence 25, Appl
22	781.5	11.8	1887	14	US-10-231-353-29	Sequence 29, Appl
23	647.5	9.8	1143	14	US-10-231-353-3	Sequence 3, Appli
24	624.5	9.5	7718	14	US-10-175-523-192	Sequence 192, App
25	607	9.2	8232	14	US-10-198-846-10976	Sequence 10976, A
26	592.5	9.0	909	14	US-10-231-353-13	Sequence 13, Appl
27	584.5	8.9	4277	11	US-09-946-374-57	Sequence 57, Appl
28	584.5	8.9	4277	12	US-10-015-387A-57	Sequence 57, Appl
29	584.5	8.9	4277	12	US-10-006-130A-57	Sequence 57, Appl
30	584.5	8.9	4277	12	US-10-199-672-439	Sequence 439, App
31	584.5	8.9	4277	12	US-10-006-172A-57	Sequence 57, Appl
32	584.5	8.9	4277	12	US-10-187-749-439	Sequence 439, App
33	584.5	8.9	4277	12	US-10-194-457-439	Sequence 439, App
34	584.5	8.9	4277	12	US-10-184-642-439	Sequence 439, App
35	584.5	8.9	4277	12	US-10-196-747-439	Sequence 439, App
36	584.5	8.9	4277	12	US-10-015-392A-57	Sequence 57, Appl
37	584.5	8.9	4277	12	US-10-017-253A-57	Sequence 57, Appl
38	584.5	8.9	4277	12	US-10-173-689-439	Sequence 439, App
39	584.5	8.9	4277	12	US-10-173-690-439	Sequence 439, App
40	584.5	8.9	4277	12	US-10-173-691-439	Sequence 439, App
41	584.5	8.9	4277	12	US-10-173-692-439	Sequence 439, App
42	584.5	8.9	4277	12	US-10-173-694-439	Sequence 439, App
43	584.5	8.9	4277	12	US-10-173-698-439	Sequence 439, App
44	584.5	8.9	4277	12	US-10-173-699-439	Sequence 439, App
45	584.5	8.9	4277	12	US-10-173-707-439	Sequence 439, App

ALIGNMENTS

```
RESULT 1
US-09-754-997A-45
; Sequence 45, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 3756
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-754-997A-45

Alignment Scores:
Pred. No.:              0             Length:        3756
```

```
Score:                  6604.00    Matches:       1252
Percent Similarity:     100.00%    Conservative:  0
Best Local Similarity:  100.00%    Mismatches:    0
Query Match:            100.00%    Indels:        0
DB:                     10         Gaps:          0

US-09-754-997A-2 (1-1252) x US-09-754-997A-45 (1-3756)

Qy      1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy     21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy     41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy     61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy     81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy    101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy    121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy    141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy    161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy    181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy    201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy    221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy    241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy    261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy    281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy    301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy    321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy    341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy    361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy    381 LysValGlnGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy    401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy    421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy    441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy    461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy    481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy    501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy    521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy    541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy    561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy    581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy    601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy    621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy    641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy    661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy    681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy        701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy        721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy        741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy        761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy        781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AACTACACTGTACGCTTCGGCCCCTGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy        801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy        821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy        841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy        861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy        881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy        901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy        921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy        941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy        961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy        981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy       1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy       1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy       1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy       1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy       1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy       1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy       1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy       1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy       1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy       1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGCCCAGAGAGGCTCACCTGCTTG 3600

Qy       1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy       1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy       1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
              ||||||||||||||||||||||||||||||||||||
Db       3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

RESULT 2
US-09-754-997A-1
; Sequence 1, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
; APPLICANT: Salbaum, Michael J.
; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
; TITLE OF INVENTION: Acids and Methods of Use
; FILE REFERENCE: P-NI 4552
; CURRENT APPLICATION NUMBER: US/09/754,997A
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,496
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: US 60/205,789
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 6176
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(3756)
US-09-754-997A-1

Alignment Scores:

Pred. No.:	0	Length:	6176
Score:	6604.00	Matches:	1252
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB:               10          Gaps:        0
US-09-754-997A-2 (1-1252) x US-09-754-997A-1 (1-6176)

Qy      1 MetAlaArgAlaAspThrGlyArgGlyLeuLeuValLeuThrPheCysLeuLeuSerAla 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCGCGGGCGGACACGGGCCGCGGGCTCCTGGTGCTGACCTTCTGCCTGCTGTCCGCG 60

Qy     21 ArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyPro 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCGGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCC 120

Qy     41 LeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGCAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACA 180

Qy     61 AlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHis 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCTGCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCAT 240

Qy     81 GluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGAACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAA 300

Qy    101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACAGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGT 360

Qy    121 LeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGGCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACA 420

Qy    141 LeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArg 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCGAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGC 480

Qy    161 PheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnVal 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTG 540

Qy    181 ThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCGTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGAT 600

Qy    201 ValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArg 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTCCAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGA 660

Qy    221 PheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGly 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCAGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGG 720

Qy    241 GlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnVal 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAGGATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTA 780

Qy    261 ValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAsp 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTGATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGAT 840

Qy    281 GlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGC 900

Qy    301 AlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAsp 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGAC 960

Qy    321 PheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 1020

Qy    341 GluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGlu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1080

Qy    361 ProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgVal 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1140

Qy    381 LysValGlnGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGly 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1200

Qy    401 TyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAla 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1260

Qy    421 ValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSer 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1320

Qy    441 SerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1380

Qy    461 PheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaVal 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1440

Qy    481 AsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPhe 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1500

Qy    501 TyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHis 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1560

Qy    521 ThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSer 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1620

Qy    541 AspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLys 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1680

Qy    561 TyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGly 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1740

Qy    581 AsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIle 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1800

Qy    601 SerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrPro 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1860

Qy    621 GlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLys 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1920

Qy    641 MetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyr 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1980

Qy    661 LysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGly 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2040

Qy    681 GlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGln 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2100

Qy    701 TyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsn 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     2101 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2160

Qy      721 LysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrPro 740
Db     2161 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2220

Qy      741 AspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsn 760
Db     2221 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2280

Qy      761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
Db     2281 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2340

Qy      781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
Db     2341 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2400

Qy      801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
Db     2401 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2460

Qy      821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
Db     2461 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2520

Qy      841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
Db     2521 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2580

Qy      861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
Db     2581 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2640

Qy      881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
Db     2641 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2700

Qy      901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
Db     2701 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2760

Qy      921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
Db     2761 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2820

Qy      941 ThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIleIleVal 960
Db     2821 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCACGCCGTCACGGGCATCATCGTG 2880

Qy      961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
Db     2881 GGTGTCTGCCTGGGCCTTCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTACGACAAAGC 2940

Qy      981 SerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeu 1000
Db     2941 TCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAACCCAGCGCTC 3000

Qy     1001 TyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGluSerLeu 1020
Db     3001 TACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTGGAGTCCCTC 3060

Qy     1021 ValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLysAlaGlu 1040
Db     3061 GTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGACAAGGCTGAA 3120

Qy     1041 ValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArgLysIle 1060
Db     3121 GTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAGAGAAAGATC 3180

Qy     1061 SerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuPro 1080
Db     3181 TCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGTGAGCTGCCC 3240

Qy     1081 GlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGlyThrGly 1100
Db     3241 CAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCGGGAACCGGG 3300

Qy     1101 GlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArgLysLys 1120
Db     3301 CAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGGAGAAAGAAG 3360

Qy     1121 ProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAspPheGly 1140
Db     3361 CCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCCGACTTCGGT 3420

Qy     1141 AlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluProLeuThr 1160
Db     3421 GCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAACCACTGACT 3480

Qy     1161 AlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAspTrpLeu 1180
Db     3481 GCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCAGACTGGCTG 3540

Qy     1181 GlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThrCysLeu 1200
Db     3541 GGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTCACCTGCTTG 3600

Qy     1201 ProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAlaIleGlu 1220
Db     3601 CCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACTGCTATAGAG 3660

Qy     1221 GluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSerProSer 1240
Db     3661 GAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTCAGCCCAAGC 3720

Qy     1241 LeuProArgAlaProValSerSerAlaGlnValPro 1252
Db     3721 CTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 3756
```

```
RESULT 3
US-10-239-663-29
; Sequence 29, Application US/10239663
; Publication No. US20030139572A1
; GENERAL INFORMATION:
;  APPLICANT: Agarwal, Pankaj
;  APPLICANT:  Murdoch, Paul R.
;  APPLICANT:  Rizvi, Safia, K.
;  APPLICANT:  Smith, Randall, F.
;  APPLICANT:  Xiang, Zhaoying
;  APPLICANT:  Kabnick, Karen
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP50018
;  CURRENT APPLICATION NUMBER: US/10/239,663
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/US01/09226
;  PRIOR FILING DATE: 2001-03-22
;  PRIOR APPLICATION NUMBER: 60/192,158
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/192,668
;  PRIOR FILING DATE: 2000-03-27
;  PRIOR APPLICATION NUMBER: 60/200,166
;  PRIOR FILING DATE: 2000-04-27
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 3753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-239-663-29

Alignment Scores:
Pred. No.:                 0             Length:         3753
Score:                     5746.00       Matches:        1098
Percent Similarity:        92.09%        Conservative:   54
Best Local Similarity:     87.77%        Mismatches:     96
```

```
Query Match:          87.01%      Indels:      5
DB:                   12          Gaps:        2

US-09-754-997A-2 (1-1252) x US-10-239-663-29 (1-3753)

Qy      1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
Db      1 ATGGCGCGGGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 60

Qy     19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
Db     61 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 120

Qy     39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
Db    121 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 180

Qy     59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
Db    181 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 240

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db    241 GAGCACGACCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 300

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db    301 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 357

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    358 TCGTGCCTAGCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 417

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    418 GCCACACTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 477

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    478 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 537

Qy    179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db    538 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 597

Qy    199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db    598 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 657

Qy    219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db    658 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACC 717

Qy    239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db    718 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 777

Qy    259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db    778 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 837

Qy    279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db    838 CAAGACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 897

Qy    299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db    898 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 957

Qy    319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGln 338
Db    958 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAG 1017

Qy    339 AlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSer 358
Db   1018 GCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCG 1077

Qy    359 GlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGly 378
Db   1078 GGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGG 1137

Qy    379 ArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAsp 398
Db   1138 CGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGAC 1197

Qy    399 AlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaPro 418
Db   1198 GCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCG 1257

Qy    419 LeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrPro 438
Db   1258 CTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCA 1317

Qy    439 LeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIle 458
Db   1318 CTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATC 1377

Qy    459 IleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPhe 478
Db   1378 ATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTT 1437

Qy    479 AlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyr 498
Db   1438 GCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTAT 1497

Qy    499 GluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeu 518
Db   1498 GAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTG 1557

Qy    519 ValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsn 538
Db   1558 GTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAAC 1617

Qy    539 ProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnVal 558
Db   1618 CCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTG 1677

Qy    559 LeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluVal 578
Db   1678 GTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTG 1737

Qy    579 ProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgVal 598
Db   1738 CGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTA 1797

Qy    599 ArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArg 618
Db   1798 CGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGG 1857

Qy    619 ThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArg 638
Db   1858 ACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAG 1917

Qy    639 AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSer 658
Db   1918 GCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCT 1977

Qy    659 GlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgPro 678
Db   1978 GGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTG 2037

Qy    679 ProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysVal 698
Db   2038 CCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTG 2097

Qy    699 LysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAla 718
```

```
Db     2098 AAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCT 2157
Qy      719 PheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaPro 738
Db     2158 TTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCG 2217
Qy      739 ThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGlu 758
Db     2218 GCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAA 2277
Qy      759 SerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLys 778
Db     2278 TCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAG 2337
Qy      779 IleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThr 798
Db     2338 ATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACC 2397
Qy      799 TyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyr 818
Db     2398 TATTACACCAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATAC 2457
Qy      819 GluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGlu 838
Db     2458 GAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAG 2517
Qy      839 ArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeu 858
Db     2518 CGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTG 2577
Qy      859 ThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleVal 878
Db     2578 ACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTG 2637
Qy      879 GluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeu 898
Db     2638 GAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTC 2697
Qy      899 ThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyr 918
Db     2698 ACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTAC 2757
Qy      919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
Db     2758 TTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGAT 2817
Qy      939 ValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaValThrGlyIle 958
Db     2818 GTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTCACGGGCATC 2877
Qy      959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
Db     2878 ATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCTGGCCTGCGC 2937
Qy      979 GlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsnPro 998
Db     2938 CGCAGCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCCGGGAATCCC 2997
Qy      999 AlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeuGlu 1018
Db     2998 GCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCATGAATTGGAG 3057
Qy     1019 SerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAspLys 1038
Db     3058 TCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTGGAGGACAGG 3117
Qy     1039 AlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLysArg 1058
Db     3118 GCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCACTCCAAAAGA 3177
Qy     1059 LysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGlu 1078
Db     3178 AAGATCTCCTGGGCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCAGGCTGTGAG 3237
Qy     1079 LeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAlaGly 1098
Db     3238 CTGCCCCAG---GCAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCCCCTGCTGGA 3294
Qy     1099 ThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGlyArg 1118
Db     3295 ACTGGGCAGACGCTGTTGCTGCAGGCTCTGGTGTACGACGCCATAAAGGGCAATGGGAGG 3354
Qy     1119 LysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSerAsp 1138
Db     3355 AAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTCCACTCTGAC 3414
Qy     1139 PheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGluPro 1158
Db     3415 TTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCTGAGGACCCC 3474
Qy     1159 LeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAlaAsp 1178
Db     3475 CTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAGGGGGCAGCC 3534
Qy     1179 TrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeuThr 1198
Db     3535 TGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGACAGACTTACC 3594
Qy     1199 CysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThrAla 1218
Db     3595 TGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCAGGCGAGGTG 3654
Qy     1219 IleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrValSer 1238
Db     3655 CTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTAGGAGCCAGC 3714
Qy     1239 ProSerLeuProArgAlaProValSerSerAla 1249
Db     3715 CCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3747
```

RESULT 4
US-09-908-193-1
; Sequence 1, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3741
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:

```
;    OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
;    OTHER INFORMATION: with homology to NOPE-like protein
;    NAME/KEY: modified_base
;    LOCATION: (375)
;    OTHER INFORMATION: a, c, t, g, other or unknown
US-09-908-193-1

Alignment Scores:
Pred. No.:              0             Length:          3741
Score:                  5610.00       Matches:         1086
Percent Similarity:     91.07%        Conservative:    56
Best Local Similarity:  86.60%        Mismatches:      101
Query Match:            84.95%        Indels:          13
DB:                     10            Gaps:            8

US-09-754-997A-2 (1-1252) x US-09-908-193-1 (1-3741)

Qy      1 MetAlaArgAlaAspThr------GlyArgGlyLeuLeuValLeuThrPheCysLeuLeu 18
Db      1 ATGGCGCGGGGGGACGCCGGCCGCGGCCGCGGGCTCCTCGCGTTGACCTTCTGCCTGTTG 60

Qy     19 SerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGlu 38
Db     61 GCCGCGCGCGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTG 120

Qy     39 GlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGly 58
Db    121 GGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGG 180

Qy     59 AlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeu 78
Db    181 GCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTG 240

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db    241 GAGCACGACCACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCA 300

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db    301 CCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGCAACTAT 357

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    358 TCGTGCCTAGCCCACGGNCCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTT 417

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    418 GCCAGTCTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACA 477

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    478 GCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGAC 537

Qy    179 GlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIle 198
Db    538 CAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATC 597

Qy    199 LeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArg 218
Db    598 CTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGC 657

Qy    219 GlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThr 238
Db    658 CAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACC 717

Qy    239 ArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGln 258
Db    718 AGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAG 777

Qy    259 AsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArg 278
Db    778 AGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA 837

Qy    279 GlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIle 298
Db    838 ---GACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATT 894

Qy    299 AlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThr 318
Db    895 GCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACG 954

Qy    319 ArgAspPheAlaThrAlaAlaAlaGluLeuArgVal---LeuAlaAlaProAlaIleSer 337
Db    955 CGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACT 1014

Qy    338 GlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAla 357
Db   1015 CAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCG 1074

Qy    358 SerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsn 377
Db   1075 TCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAAC 1134

Qy    378 GlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGln 397
Db   1135 GGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAG 1194

Qy    398 AspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAla 417
Db   1195 GACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCG 1254

Qy    418 ProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThr 437
Db   1255 TCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACG 1314

Qy    438 ProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGln 457
Db   1315 CCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAG 1374

Qy    458 IleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGln 477
Db   1375 ATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAG 1434

Qy    478 PheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAsp 497
Db   1435 TTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGAT 1494

Qy    498 TyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAla 517
Db   1495 TATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCA 1554

Qy    518 LeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerPro 537
Db   1555 CTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCC 1614

Qy    538 AsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGln 557
Db   1615 AACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAG 1674

Qy    558 ValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGlu 577
Db   1675 GTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAG 1734

Qy    578 ValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArg 597
Db   1735 GTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGA 1794

Qy    598 ValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHis 617
Db   1795 GTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCAC 1854

Qy    618 ArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysVal 637
Db   1855 AGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTG 1914

Qy    638 ArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGlnIle 657
```

```
Db     1915 CAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATC 1974

Qy      658 SerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArg 677
Db     1975 TCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGC 2034

Qy      678 ProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLys 697
Db     2035 CTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAA 2094

Qy      698 ValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuVal 717
Db     2095 GTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTG 2154

Qy      718 AlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAla 737
Db     2155 GCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCG 2214

Qy      738 ProThrProAspLeuProIleGlnArgGlyProProLeuProProAlaHisValHisAla 757
Db     2215 CCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCG 2274

Qy      758 GluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrVal 777
Db     2275 GAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTC 2334

Qy      778 LysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuVal 797
Db     2335 AAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTC 2394

Qy      798 ThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLys 817
Db     2395 ACCTATTAC---AGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAA 2451

Qy      818 TyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValVal 837
Db     2452 TACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTG 2511

Qy      838 GluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerPro 857
Db     2512 GAGCGCTCCACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCC 2571

Qy      858 LeuThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIle 877
Db     2572 CTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATC 2631

Qy      878 ValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeu 897
Db     2632 GTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTG 2691

Qy      898 LeuThrThr------GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAsp 915
Db     2692 CTCACCACGCAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGAC 2751

Qy      916 ThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArg 935
Db     2752 ACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGC 2811

Qy      936 LeuGlnAspValIleThrLeuGlnGluThrPheSerAspSerLeuAspValHisAlaVal 955
Db     2812 CTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCACTCAGTC 2871

Qy      956 ThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAla 975
Db     2872 ACGGGCATCATCGTGGGTGTCTGCCTGGGCCTCCTCTGCCTCCTGGCCTGCATGTGTGCT 2931

Qy      976 GlyLeuArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerGlyThrPro 995
Db     2932 GGCCTGCGCCGCAGCCCCCACAGGGAATCCCTCCCAGGCCTGTCCTCCACCGCCACCCCC 2991

Qy      996 GlyAsnProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHis 1015
Db     2992 GGGAATCCCGCGCTGTACTCCAGAGCTCGGCTTGGCCCCCCCAGCCCCCCAGCTGCCCAT 3051

Qy     1016 GluLeuGluSerLeuValHisProArgProGlnAspTrpSerProProProSerAspVal 1035
Db     3052 GAATTGGAGTCCCTTGTGCACCCCCATCCCCAGGACTGGTCCCCGCCACCCTCAGACGTG 3111

Qy     1036 GluAspLysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHis 1055
Db     3112 GAGGACAGGGCTGAAGTGCACAGCCTTATGGGTGGCGGTGTTTCTGAAGGCCGGAGTCAC 3171

Qy     1056 SerLysArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAla 1075
Db     3172 TCCAAAAGAAAGGTAAGT---GCTCAACCAAGCGGGCTGAGCTGGGCTGGTTCCTGGGCA 3228

Qy     1076 GlyCysGluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuPro 1095
Db     3229 GGCTGTGAGCTGCCCCAG---GCAGGCCCCCGGCCGGCTCTGACCCGGGCCCTGCTGCCC 3285

Qy     1096 ProAlaGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSer 1115
Db     3286 CCTGCTGGAACTGGGCAGACGCTGTTGCTGCAGGTTCTCTGCTCTGAT------CAGGGC 3339

Qy     1116 AsnGlyArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleVal 1135
Db     3340 AATGGGAGGAAGAAGTCACCCCCAGCCTGCAGGAACCAGGTGGAGGCTGAAGTCATTGTC 3399

Qy     1136 HisSerAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluPro 1155
Db     3400 CACTCTGACTTTAGTGCATCTAACGGGAACCCTGACCTCCATCTCCAAGACCTGGAGCCT 3459

Qy     1156 GluGluProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGln 1175
Db     3460 GAGGACCCCCTGCCTCCAGAGGCTCCTGATCTCATCTCGGGTGTTGGGGATCCAGGGCAG 3519

Qy     1176 GlyAlaAspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGlu 1195
Db     3520 GGGGCAGCCTGGCTGGACAGGGAGTTGGGAGGGTGTGAGCTGGCAGCCCCCGGGCCAGAC 3579

Qy     1196 ArgLeuThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnPro 1215
Db     3580 AGACTTACCTGCTTGCCAGAGGCAGCCAGTGCTTCCTGCTCCTACCCGGACCTCCAGCCA 3639

Qy     1216 SerThrAlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeu 1235
Db     3640 GGCGAGGTGCTAGAGGAGACCCCTGGAGATAGCTGCCAGCTCAAATCCCCCTGCCCTCTA 3699

Qy     1236 ThrValSerProSerLeuProArgAlaProValSerSerAla 1249
Db     3700 GGAGCCAGCCCAGGCCTGCCCAGATCCCCGGTCTCCTCCTCT 3741

RESULT 5
US-09-754-997A-3
; Sequence 3, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
;  APPLICANT: Salbaum, Michael J.
;  TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
;  TITLE OF INVENTION:  Acids and Methods of Use
;  FILE REFERENCE: P-NI 4552
;  CURRENT APPLICATION NUMBER: US/09/754,997A
;  CURRENT FILING DATE:  2001-01-04
;  PRIOR APPLICATION NUMBER: US 60/174,496
;  PRIOR FILING DATE: 2000-01-04
;  PRIOR APPLICATION NUMBER: US 60/205,789
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 2796
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
```

```
;   LOCATION: (1)...(2796)
US-09-754-997A-3

Alignment Scores:
Pred. No.:               0           Length:         2796
Score:                   4903.00     Matches:        932
Percent Similarity:      100.00%     Conservative:   0
Best Local Similarity:   100.00%     Mismatches:     0
Query Match:             74.24%      Indels:         0
DB:                      10          Gaps:           0

US-09-754-997A-2 (1-1252) x US-09-754-997A-3 (1-2796)

Qy       22 GlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCysAspGluGlyProLeu 41
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy       42 GlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAla 61
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy       62 AlaGlyProProThrArgValThrTrpSerLysAspGlyAspThrValLeuGluHisGlu 81
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy       82 AsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAsp 101
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy      102 SerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCysLeu 121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy      122 AlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAlaThrLeu 141
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy      142 GluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAlaArgPhe 161
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy      162 GluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGlnValThr 181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy      182 ValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspVal 201
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy      202 GlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPhe 221
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy      222 SerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGln 241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy      242 AspValValIleValAlaAlaProGluAsnThrThrValValSerGlyGlnAsnValVal 261
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy      262 MetGluCysValAlaSerAlaAspProThrProPheValSerTrpValArgGlnAspGly 281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy      282 LysProIleSerThrAspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAla 301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy      302 GlnProArgHisSerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPhe 321
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy      322 AlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGlu 341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy      342 AlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluPro 361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy      362 ArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLys 381
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy      382 ValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyr 401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy      402 TyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaVal 421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy      422 ValValArgGluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSer 441
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy      442 SerSerValLeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPhe 461
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy      462 SerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsn 481
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy      482 AsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyr 501
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy      502 ValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThr 521
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy      522 LeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAsp 541
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy      542 IleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyr 561
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy      562 LysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsn 581
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Qy      582 GluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSer 601
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy      602 AlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGly 621
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy      622 ValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMet 641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy      642 GluSerLeuValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLys 661
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy      662 LeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGly 681
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy      682 ArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyr 701
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy        702 GluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLys 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy        722 HisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAsp 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy        742 LeuProIleGlnArgGlyProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy        762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy        782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy        802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy        822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy        842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy        862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy        882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy        902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy        942 LeuGlnGluThrPheSerAspSerLeuAspValHis 953
              ||||||||||||||||||||||||||||||||||||
Db       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

RESULT 6
US-10-239-663-28
; Sequence 28, Application US/10239663
; Publication No. US20030139572A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kabnick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50018
; CURRENT APPLICATION NUMBER: US/10/239,663
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-239-663-28

```
Alignment Scores:
Pred. No.:                 2.34e-246      Length:          2223
Score:                     3167.50        Matches:         620
Percent Similarity:        90.33%         Conservative:    34
Best Local Similarity:     85.64%         Mismatches:      54
Query Match:               47.96%         Indels:          18
DB:                        12             Gaps:            2
```

US-09-754-997A-2 (1-1252) x US-10-239-663-28 (1-2223)

```
Qy         17 LeuLeuSerAlaArgGlyGluLeuProLeuProGlnGluThrThrValLysLeuSerCys 36
              ||||||           |||||||||  ||||||||||||||||||||:::||||||||
Db         61 CTCCTCCTCTATGCAGGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGT 120

Qy         37 AspGluGlyProLeuGlnValIleLeuGlyProGluGlnAlaValValLeuAspCysThr 56
                    ||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db        121 GGAGTGGGGCCACTGCAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGC 180

Qy         57 LeuGlyAlaThrAlaAlaGlyProProThrArgValThrTrpSerLysAspGlyAspThr 76
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGGGGCTGCTGCCGCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACC 240

Qy         77 ValLeuGluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerPro 96
              :::|||||||||::::::|||||||||||||||||||||||||||||||||||   |||
Db        241 CTGCTGGAGCACGACCACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCA 300

Qy         97 LeuGluGlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGly 116
              |||         ||||||:::   ::::::||||||:::|||   |||   ||||||
Db        301 CTAGCACCCAATGGCAGTGACGAGT--CAGTCCCTGAGGCTGTGGGG-GTCATTGAAGGC 357

Qy        117 SerTyrSerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValVal 136
              :::||||||||||||||||||   |||   ||||||:::|||||||||   |||||||||
Db        358 AACTATTCGTGCCTAGCCCACGGCCCCCCTGGAGTGCTGGCCAGCCAGACTGCTGTCGTC 417

Qy        137 LysLeuAlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsn 156
              |||||||||||||   |||||||||||||||||||||||||||||   ||||||||||||
Db        418 AAGCTTGCCACACTCGCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAAC 477

Qy        157 GlyThrAlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGlu 176
              ||||||||||||||||||||||||   :::||||||||||||||||||||||||||||||
Db        478 GGGACAGCTCGCTTTGAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAG 537

Qy        177 LysAspGlnValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeu 196
              |||||||||||||||:::||||||||||||||||||||||   ||||||      :::|||
Db        538 AAGGACCAGGTGACATTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTT 597

Qy        197 GlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSer 216
              ||||||||||||||||||:::|||||||||||||   ||||||||||||||||||||||||
Db        598 CAGATCCTGGATGTTCAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCA 657

Qy        217 AlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGlu 236
              |||||||||   |||||||||||||||   |||:::||||||   ||||||||||||
Db        658 GCTCGCCAGCACTTCAGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCG 717

Qy        237 AlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSer 256
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 TCCACCAGGGGGCAGGACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCT 777

Qy        257 GlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrp 276
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    778 GGCCAGAGTGTGGTGATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGG 837

Qy    277 ValArgGlnAspGlyLysProIleSerThrAspValIleValLeuGlyArgThrAsnLeu 296
Db    838 GTCCGACAAGACGGGAAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTA 897

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    898 CTAATTGCCAACGCGCAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCC 957

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    958 CGCACGCGCGACTTCGCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATC 1017

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db   1018 ACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGC 1077

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db   1078 GCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCC 1137

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db   1138 AACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTG 1197

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db   1198 CAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCC 1257

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db   1258 GCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCT 1317

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1318 ACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAG 1377

Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1378 CAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATAC 1437

Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1438 CAGTTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACA 1497

Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1498 GATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCA 1557

Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1558 GCACTGGTGCACACACTGGATGAT------GGTAGGGCCTCTGAACTCGCAGTGGGCAGC 1611

Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1612 -------------------------------------------TTGGGCCTGAGCAATGGG 1629

Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1630 CAGGTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACT 1689

Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1690 GAGGTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTAT 1749

Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1750 CGAGTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCAT 1809

Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1810 CACAGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAG 1869

Qy    637 ValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGln 656
Db   1870 GTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAG 1929

Qy    657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
Db   1930 ATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGAT 1989

Qy    677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
Db   1990 CGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAG 2049

Qy    697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
Db   2050 AAAGTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTC 2109

Qy    717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
Db   2110 GTGGCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAG 2169

Qy    737 AlaProThrPro 740
Db   2170 GCGCCGGCACCA 2181
```

```
RESULT 7
US-10-231-353-7
; Sequence 7, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides E
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/10/231,353
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/877,730
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3453
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-7

Alignment Scores:
Pred. No.:               2.05e-131      Length:         3453
Score:                   1747.50        Matches:        397
Percent Similarity:      54.95%         Conservative:   169
Best Local Similarity:   38.54%         Mismatches:     381
Query Match:             26.46%         Indels:         83
DB:                      14             Gaps:           18

US-09-754-997A-2 (1-1252) x US-10-231-353-7 (1-3453)

Qy      9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db     43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy     29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db     97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy     48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db    157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207
```

```
Qy     68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db    208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267
Qy     88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db    268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318
Qy    108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db    319 ------------------GATGAAGGATTTATCAGTGCTTGGCAATGAACAAATATGGA 360
Qy    128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db    361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420
Qy    148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db    421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480
Qy    168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db    481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540
Qy    187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db    541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600
Qy    207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db    601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660
Qy    227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCAACAATTATA 714
Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTGGAATGCATGGCC 774
Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834
Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954
Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014
Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074
Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128
Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188
Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248
Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308
Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368
Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548
Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668
Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728
Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788
Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845
Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902
Qy    645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962
Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   1963 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 1989
Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   1990 ------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031
Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091
Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2092 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136
Qy    744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196
Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256
Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
```

```
Db   2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316

Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376

Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436

Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496

Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556

Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616

Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676

Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736

Qy    940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db   2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796

Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db   2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856

Qy    960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db   2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916

Qy    980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db   2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976

Qy    988 GlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006
```

RESULT 8
US-10-231-353-1
; Sequence 1, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Enc
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3210
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-231-353-1

```
Alignment Scores:
Pred. No.:                2.25e-131     Length:        3210
Score:                    1746.50       Matches:       397
Percent Similarity:       54.95%        Conservative:  169
Best Local Similarity:    38.54%        Mismatches:    381
Query Match:              26.45%        Indels:        83
DB:                       14            Gaps:          18

US-09-754-997A-2 (1-1252) x US-10-231-353-1 (1-3210)

Qy      9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db     43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy     29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db     97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy     48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db    157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy     68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db    208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy     88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db    268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy    108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db    319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy    128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db    361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy    148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db    421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy    168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db    481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy    187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db    541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy    207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db    601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy    227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
```

```
Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902

Qy    645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962

Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   1963 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 1989

Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   1990 -----------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031

Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091

Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2092 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136

Qy    744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196

Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256

Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db   2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316

Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376

Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436

Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496

Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556

Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616

Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676

Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736

Qy    940 IleThrLeuGlnGluThrPheSerAspSer----------------------------- 949
Db   2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796

Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db   2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856

Qy    960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db   2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916

Qy    980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db   2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976

Qy    988 GlyLeuSerSerSerGlyThrProGlyAsn 997
```

```
Db       2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006

RESULT 9
US-10-231-353-31
; Sequence 31, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Encc
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/10/231,353
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/877,730
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 3874
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-31

Alignment Scores:
Pred. No.:              2.84e-131     Length:        3874
Score:                  1746.50       Matches:       397
Percent Similarity:     54.95%        Conservative:  169
Best Local Similarity:  38.54%        Mismatches:    381
Query Match:            26.45%        Indels:        83
DB:                     14            Gaps:          18

US-09-754-997A-2 (1-1252) x US-10-231-353-31 (1-3874)

Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db        233 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 286

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db        287 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 346

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db        347 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 397

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db        398 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 457

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db        458 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 508

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db        509 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 550

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db        551 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 610

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db        611 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 670

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db        671 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 730

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db        731 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 790

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db        791 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 850

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db        851 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 904

Qy        247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db        905 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 964

Qy        267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db        965 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 1024

Qy        287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db       1025 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 1084

Qy        306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db       1085 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 1144

Qy        326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db       1145 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1204

Qy        346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db       1205 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1264

Qy        366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db       1265 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1318

Qy        386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db       1319 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1378

Qy        406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaProLeuAlaValValValArgGlu 425
Db       1379 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1438

Qy        426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db       1439 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1498

Qy        446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db       1499 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1558

Qy        466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db       1559 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1618

Qy        486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db       1619 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1678

Qy        506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db       1679 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1738

Qy        526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db       1739 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1798
```

```
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1799 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1858

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1859 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1918

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1919 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1978

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1979 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 2035

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   2036 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 2092

Qy    645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   2093 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 2152

Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   2153 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 2179

Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   2180 ------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2221

Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2222 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2281

Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2282 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2326

Qy    744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2327 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2386

Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2387 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2446

Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db   2447 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2506

Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2507 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2566

Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2567 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2626

Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2627 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2686

Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2687 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2746

Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2747 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2806

Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2807 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2866

Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2867 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2926

Qy    940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db   2927 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2986

Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db   2987 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 3046

Qy    960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db   3047 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 3106

Qy    980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db   3107 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 3166

Qy    988 GlyLeuSerSerSerGlyThrProGlyAsn 997
Db   3167 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3196

RESULT 10
US-10-231-353-11
; Sequence 11, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides En
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/10/231,353
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: US/09/877,730
;  PRIOR FILING DATE: 2001-06-08
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 2976
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-11

Alignment Scores:
Pred. No.:                5.68e-127      Length:         2976
Score:                    1691.50        Matches:        375
Percent Similarity:       55.79%         Conservative:   160
Best Local Similarity:    39.10%         Mismatches:     347
Query Match:              25.61%         Indels:         77
DB:                       14             Gaps:           15

US-09-754-997A-2 (1-1252) x US-10-231-353-11 (1-2976)

Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
```

```
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159
Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219
Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGAGTTCAAT 279
Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339
Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399
Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453
Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513
Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573
Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633
Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693
Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753
Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813
Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873
Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927
Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987
Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047
Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107
Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167
Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227

Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGAGCCAGCCAGATGTCTGACCAT 1287
Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347
Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGC 1407
Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467
Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527
Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587
Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1588 CATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641
Qy    637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
Db   1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701
Qy    656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAsp 674
Db   1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755
Qy    675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694
Db   1756 ------------------------------------------GGGCCCATTTTCTTG 1770
Qy    695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714
Db   1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830
Qy    715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734
Db   1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875
Qy    735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750
Db   1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935
Qy    751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770
Db   1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995
Qy    771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790
Db   1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055
Qy    791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810
Db   2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTCAA 2115
Qy    811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830
Db   2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175
Qy    831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
Db   2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235
```

```
Qy    851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
Db   2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295

Qy    871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
Db   2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355

Qy    891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
Db   2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415

Qy    911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
Db   2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475

Qy    931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949
Db   2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535

Qy    950 ---------------------------------------------------------Leu 950
Db   2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595

Qy    951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
Db   2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655

Qy    971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985
Db   2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715

Qy    986 ---------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 11
US-10-231-353-17
; Sequence 17, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064Alel Human Phosphatases and Polynucleotides Enco
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 3219
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-231-353-17

Alignment Scores:

Pred. No.:	6.26e-127	Length:	3219
Score:	1691.50	Matches:	375
Percent Similarity:	55.79%	Conservative:	160
Best Local Similarity:	39.10%	Mismatches:	347
Query Match:	25.61%	Indels:	77
DB:	14	Gaps:	15

US-09-754-997A-2 (1-1252) x US-10-231-353-17 (1-3219)

```
Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
```

```
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107
Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167
Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227
Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287
Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347
Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGC 1407
Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467
Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527
Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587
Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1588 CATAGGACGCCCAAAGCT--ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641
Qy    637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
Db   1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701
Qy    656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAsp 674
Db   1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755
Qy    675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694
Db   1756 -------------------------------------------GGGCCCATTTTCTTG 1770
Qy    695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714
Db   1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830
Qy    715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734
Db   1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875
Qy    735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750
Db   1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935
Qy    751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770
Db   1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995
Qy    771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790
Db   1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055
Qy    791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810
Db   2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAA 2115
Qy    811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830
Db   2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175
Qy    831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
Db   2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235
Qy    851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
Db   2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295
Qy    871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
Db   2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355
Qy    891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
Db   2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415
Qy    911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
Db   2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475
Qy    931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949
Db   2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535
Qy    950 ---------------------------------------------------------Leu 950
Db   2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595
Qy    951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
Db   2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655
Qy    971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985
Db   2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715
Qy    986 ---------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 12
US-09-754-997A-5
; Sequence 5, Application US/09754997A
; Patent No. US20020102551A1
; GENERAL INFORMATION:
; APPLICANT: Salbaum, Michael J.
; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
; TITLE OF INVENTION: Acids and Methods of Use
; FILE REFERENCE: P-NI 4552
; CURRENT APPLICATION NUMBER: US/09/754,997A
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,496
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: US 60/205,789
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(825)
US-09-754-997A-5

Alignment Scores:
Pred. No.: 6.19e-110 Length: 825

```
Score:                    1472.00      Matches:        275
Percent Similarity:       100.00%      Conservative:   0
Best Local Similarity:    100.00%      Mismatches:     0
Query Match:              22.29%       Indels:         0
DB:                       10           Gaps:           0

US-09-754-997A-2 (1-1252) x US-09-754-997A-5 (1-825)

Qy        978 ArgGlnSerSerHisArgGluAlaLeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGACAAAGCTCCCACAGGGAAGCCCTTCCCGGATTGTCCTCCTCAGGCACCCCAGGAAAC 60

Qy        998 ProAlaLeuTyrThrArgAlaArgLeuGlyProProSerValProAlaAlaHisGluLeu 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGCGCTCTACACAAGAGCTCGGCTTGGGCCTCCCAGTGTCCCTGCTGCCCATGAGTTG 120

Qy       1018 GluSerLeuValHisProArgProGlnAspTrpSerProProProSerAspValGluAsp 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGTCCCTCGTGCATCCTCGTCCCCAGGATTGGTCCCCACCACCCTCAGATGTGGAAGAC 180

Qy       1038 LysAlaGluValHisSerLeuMetGlyGlySerValSerAspCysArgGlyHisSerLys 1057
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAGGCTGAAGTACACAGCCTTATGGGTGGCAGTGTTTCAGATTGCCGGGGCCACTCCAAG 240

Qy       1058 ArgLysIleSerTrpAlaGlnAlaGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCys 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAGATCTCCTGGGCTCAGGCAGGGGGACCAAACTGGGCAGGCTCCTGGGCAGGCTGT 300

Qy       1078 GluLeuProGlnGlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAla 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGCTGCCCCAGGGTAGTGGTCCAAGGCCGGCTCTGACCCGTGCTCTGCTGCCTCCAGCG 360

Qy       1098 GlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGly 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGAACCGGGCAGACACTGCTGCTGCAAGCCCTGGTGTATGACGGCATAAAGAGCAACGGG 420

Qy       1118 ArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSer 1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAGAAGCCGTCCCCAGCCTGCAGGAATCAGGTGGAAGCTGAGGTCATTGTCCACTCC 480

Qy       1138 AspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGlu 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTCGGTGCATCCAAAGGATGTCCTGACCTCCACCTCCAAGACCTGGAGCCAGAGGAA 540

Qy       1158 ProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAla 1177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCACTGACTGCAGAGACTCTGCCTTCCACGTCTGGAGCTGTGGATCTGTCTCAAGGAGCA 600

Qy       1178 AspTrpLeuGlyArgGluLeuGlyGlyCysGlnProThrThrSerGlyProGluArgLeu 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTGGCTGGGCAGGGAGCTGGGAGGGTGCCAACCAACAACCAGTGGGCCAGAGAGGCTC 660

Qy       1198 ThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnProSerThr 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCTGCTTGCCAGAAGCAGCCAGTGCCTCCTGCTCCTGCTCAGACCTCCAGCCCAGCACT 720

Qy       1218 AlaIleGluGluAlaProGlyLysSerCysGlnProLysAlaLeuCysProLeuThrVal 1237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTATAGAGGAGGCCCCTGGGAAAAGCTGCCAGCCCAAAGCCCTGTGTCCTCTAACAGTC 780

Qy       1238 SerProSerLeuProArgAlaProValSerSerAlaGlnValPro 1252
              |||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGCCCAAGCCTTCCCAGGGCCCCTGTCTCCTCTGCTCAGGTCCCC 825
```

RESULT 13
US-10-231-353-5
; Sequence 5, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides En
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2715
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-231-353-5

```
Alignment Scores:
Pred. No.:               3.49e-102     Length:         2715
Score:                   1384.00       Matches:        334
Percent Similarity:      46.11%        Conservative:   140
Best Local Similarity:   32.49%        Mismatches:     310
Query Match:             20.96%        Indels:         244
DB:                      14            Gaps:           14

US-09-754-997A-2 (1-1252) x US-10-231-353-5 (1-2715)

Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
               |||:::|||:::                          |||||      |||||||||
Db         43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                         :::|||||     :::    |||       |||    :::
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
               :::        ||||||||||||||                        |||   :::
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
               |||||||||   |||:::|||       :::   |||::::::   :::   :::|||
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
               ||||||||||||:::::::|||               |||    |||        |||::::::
Db        268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                  ||||||   |||   ||||||||||    :::      |||
Db        319 -----------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::   ||||||   |||   :::   |||:::|||:::         |||   :::
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
               |||   |||         |||   |||   |||   |||||||||   |||
Db        421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
                  |||      :::|||||||||||||   ::::::   |||:::|||
Db        481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
               |||:::         ||||||   |||||||||||   ||||||         |||:::
Db        541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
               |||:::|||||||||:::|||         |||   :::|||   |||   |||||||||
Db        601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
               |||||||||         :::   |||:::                        |||:::
```

```
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
Db   1554 ------------------------------------------------------------ 1554

Qy    646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   1554 ------------------------------------------------------------ 1554

Qy    666 GluValGlyThrGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   1554 ------------------------------------------------------------ 1554

Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   1554 ------------------------------------------------------------ 1554

Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602

Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   1603 TATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647

Qy    746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db   1648 CGTGATCGCATGGTCCCTCCTCCACCACCCCCACCATCTCTATGCGAAGGCTAACACC 1707

Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db   1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767

Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db   1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827

Qy    802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db   1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887

Qy    822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db   1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947

Qy    842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db   1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007

Qy    862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
Db   2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067

Qy    882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db   2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127

Qy    902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
Db   2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187

Qy    922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
Db   2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247

Qy    942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
Db   2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307
```

```
Qy       950 -----------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
Db      2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367

Qy       962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
Db      2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427

Qy       982 HisArgGluAla--------------------------------LeuProGlyLeu 989
Db      2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487

Qy       990 SerSerSerGlyThrProGlyAsn 997
Db      2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

RESULT 14
US-10-231-353-9
; Sequence 9, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Enco
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 2958
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-231-353-9

Alignment Scores:

Pred. No.:	3.89e-102	Length:	2958
Score:	1384.00	Matches:	334
Percent Similarity:	46.11%	Conservative:	140
Best Local Similarity:	32.49%	Mismatches:	310
Query Match:	20.96%	Indels:	244
DB:	14	Gaps:	14

US-09-754-997A-2 (1-1252) x US-10-231-353-9 (1-2958)

```
Qy         9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db        43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy        29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db        97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy        48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db       157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy        68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db       208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy        88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db       268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy       108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db       319 -----------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy       128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db       361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy       148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db       421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy       168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db       481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy       187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db       541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy       207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db       601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy       227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db       661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy       247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db       715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy       267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db       775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy       287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db       835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy       306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db       895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy       326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db       955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy       346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db      1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy       366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db      1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy       386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db      1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy       406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db      1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy       426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db      1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy       446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db      1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy       466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
```

```
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548
Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554
Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554
Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554
Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554
Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
Db   1554 ------------------------------------------------------------ 1554
Qy    646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   1554 ------------------------------------------------------------ 1554
Qy    666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   1554 ------------------------------------------------------------ 1554
Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   1554 ------------------------------------------------------------ 1554
Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602
Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   1603 TATCAGGCA-----------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647
Qy    746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db   1648 CGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACC 1707
Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db   1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767
Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db   1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827
Qy    802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db   1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887
Qy    822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db   1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947
Qy    842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db   1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007
Qy    862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
Db   2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067
Qy    882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db   2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127
Qy    902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
Db   2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187
Qy    922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
Db   2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247
Qy    942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
Db   2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307
Qy    950 ------------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
Db   2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367
Qy    962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
Db   2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427
Qy    982 HisArgGluAla------------------------------------LeuProGlyLeu 989
Db   2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487
Qy    990 SerSerSerGlyThrProGlyAsn 997
Db   2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

RESULT 15
US-10-231-353-15
; Sequence 15, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064Alel Human Phosphatases and Polynucleotides En
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-231-353-15

Alignment Scores:

Pred. No.:	8.66e-98	Length:	2481
Score:	1329.00	Matches:	312
Percent Similarity:	46.29%	Conservative:	131
Best Local Similarity:	32.60%	Mismatches:	276
Query Match:	20.12%	Indels:	238

```
DB:                 14          Gaps:       11
US-09-754-997A-2 (1-1252) x US-10-231-353-15 (1-2481)

Qy      79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db       7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy      99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db      58 GTGGAAGGCAGGCGAGGAGAGCAGTCC-----------------GATGAAGGATTTTAT 99

Qy     119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db     100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy     139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db     160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy     159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db     220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy     179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db     280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy     198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db     340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy     218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db     400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy     238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db     454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy     258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db     514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy     278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db     574 CGCCTTGATCACAAATCCATTGATGTCTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy     297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db     634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy     317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db     694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy     337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db     754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy     357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db     814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy     377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db     874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy     397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db     928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy     417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db     988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy     437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db    1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107

Qy     457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db    1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167

Qy     477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db    1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227

Qy     497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db    1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287

Qy     517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db    1288 GTGACACAGAATACTCTAGAGGATGACCCCAGA------------------------- 1320

Qy     537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db    1320 ---------------------------------------------------------- 1320

Qy     557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db    1320 ---------------------------------------------------------- 1320

Qy     577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db    1320 ---------------------------------------------------------- 1320

Qy     597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db    1320 ---------------------------------------------------------- 1320

Qy     617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db    1320 ---------------------------------------------------------- 1320

Qy     637 ValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGln 656
Db    1320 ---------------------------------------------------------- 1320

Qy     657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
Db    1320 ---------------------------------------------------------- 1320

Qy     677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
Db    1320 ---------------------------------------------------------- 1320

Qy     697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
Db    1321 ------------------------------------AGAAAATATCATGTGAGACTC 1341

Qy     717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
Db    1342 CTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GATCAGACT 1386

Qy     737 AlaProThrProAspLeuProIleGlnArgGly------------ProProLeuProPro 752
Db    1387 GTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCC 1446

Qy     753 AlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysPro 772
Db    1447 CACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCT 1506

Qy     773 AspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArg 792
```

```
Db       1507 GCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAG 1566

Qy        793 AsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeu 812
              ||||||||||||||    |||        :::|||          :::|||:::   ||||||
Db       1567 AATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTA 1626

Qy        813 LysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyPro 832
              :::|||   ||||||||||||||||||||:::   |||           :::      |||
Db       1627 GAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCT 1686

Qy        833 PheGlySerValValGluArgSerThrLeuProAspArgProSerThrProProSerAsp 852
              :::      ||||||   |   ||||||||||||:::   |||:::   ||||||
Db       1687 TGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGA 1746

Qy        853 LeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGlu 872
              ::::::::::::   :::             |||      :::   |||   ||||||
Db       1747 GTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGC 1806

Qy        873 ProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGlu 892
              |||         :::      |   |||   |||||||||::::::
Db       1807 CCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCA 1866

Qy        893 HisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeu 912
                 :::|||   :::|||  |     ||||||   |||      |||   :::      |||
Db       1867 GGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTG 1926

Qy        913 GluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyPro 932
                 :::         |||  |     |||:::   |||      |||||||||   ||||||
Db       1927 GTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCC 1986

Qy        933 PheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--------- 949
              ||||||         ::::::         |||   :::      |||:::|||
Db       1987 TTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGG 2046

Qy        950 -------------------------------------------------LeuAspVal 952
                                                               ||||||
Db       2047 CCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAA 2106

Qy        953 HisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCys 972
                 ::::::|||||||||   |||||||||      :::   |||   |||:::|||   |||
Db       2107 AAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGT 2166

Qy        973 MetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------------- 985
              :::         :::   :::|||      |||::::::
Db       2167 GTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAG 2226

Qy        986 ---------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
                             ||||||         |||:::|||         ||||||
Db       2227 AATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2277
```

Search completed: October 14, 2003, 09:45:00
Job time : 1026 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         October 14, 2003, 06:11:10 ; Search time 137 Seconds
                                          (without alignments)
                                          4033.662 Million cell updates/sec

Title:          US-09-754-997A-2
Perfect score:  6604
Sequence:       1 MARADTGRGLLVLTFCLLSA..........CPLTVSPSLPRAPVSSAQVP 1252

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09754997/runat_14102003_061101_9736/app_query.fasta_1.1415
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09754997_@CGN_1_1_52_@runat_14102003_061101_9736 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1747.5	26.5	3453	4	US-09-877-730-7	Sequence 7, Appli
2	1746.5	26.4	3210	4	US-09-877-730-1	Sequence 1, Appli
3	1746.5	26.4	3874	4	US-09-877-730-31	Sequence 31, Appl
4	1691.5	25.6	2976	4	US-09-877-730-11	Sequence 11, Appl
5	1691.5	25.6	3219	4	US-09-877-730-17	Sequence 17, Appl
6	1384	21.0	2715	4	US-09-877-730-5	Sequence 5, Appli
7	1384	21.0	2958	4	US-09-877-730-9	Sequence 9, Appli
8	1329	20.1	2481	4	US-09-877-730-15	Sequence 15, Appl
9	1329	20.1	2724	4	US-09-877-730-19	Sequence 19, Appl
10	1144	17.3	2139	4	US-09-877-730-21	Sequence 21, Appl
11	1144	17.3	2382	4	US-09-877-730-27	Sequence 27, Appl
12	943	14.3	4608	3	US-09-041-886-24	Sequence 24, Appl
13	943	14.3	4608	5	PCT-US94-05277-1	Sequence 1, Appli
14	909	13.8	1875	4	US-09-877-730-23	Sequence 23, Appl
15	781.5	11.8	1644	4	US-09-877-730-25	Sequence 25, Appl
16	781.5	11.8	1887	4	US-09-877-730-29	Sequence 29, Appl
17	647.5	9.8	1143	4	US-09-877-730-3	Sequence 3, Appli
18	638	9.7	6000	1	US-08-348-006B-6	Sequence 6, Appli
19	638	9.7	6000	2	US-08-800-825A-6	Sequence 6, Appli
20	638	9.7	6000	3	US-09-158-657-6	Sequence 6, Appli
21	638	9.7	6000	5	PCT-US94-10166-6	Sequence 6, Appli
22	634	9.6	4078	4	US-09-016-434-1132	Sequence 1132, Ap
23	592.5	9.0	909	4	US-09-877-730-13	Sequence 13, Appl
24	548.5	8.3	3888	4	US-08-506-296B-13	Sequence 13, Appl
25	540.5	8.2	3943	4	US-08-506-296B-27	Sequence 27, Appl
26	533	8.1	3774	2	US-08-341-843B-1	Sequence 1, Appli
27	533	8.1	3774	2	US-08-427-497B-1	Sequence 1, Appli
28	533	8.1	3774	2	US-08-427-497B-2	Sequence 2, Appli
29	522	7.9	5690	2	US-08-447-464-2	Sequence 2, Appli
30	522	7.9	5690	2	US-08-716-679-2	Sequence 2, Appli
31	514	7.8	3783	4	US-08-506-296B-20	Sequence 20, Appl
32	500.5	7.6	3189	2	US-08-427-497B-3	Sequence 3, Appli
33	500.5	7.6	5824	4	US-09-620-312D-72	Sequence 72, Appl
34	477	7.2	3991	4	US-08-506-296B-3	Sequence 3, Appli
35	443	6.7	6814	4	US-09-484-970B-66	Sequence 66, Appl
36	436.5	6.6	3360	1	US-08-408-093-5	Sequence 5, Appli
37	436.5	6.6	3360	1	US-08-408-420A-5	Sequence 5, Appli
38	436.5	6.6	3360	1	US-08-714-901-5	Sequence 5, Appli
39	436.5	6.6	3360	3	US-08-040-741-5	Sequence 5, Appli
40	387	5.9	2600	2	US-08-427-497B-4	Sequence 4, Appli
41	349.5	5.3	13857	4	US-09-620-312D-75	Sequence 75, Appl
42	324.5	4.9	3024	1	US-07-923-976-7	Sequence 7, Appli
43	320	4.8	4078	4	US-09-016-434-1120	Sequence 1120, Ap
44	320	4.8	4724	1	US-08-404-665-3	Sequence 3, Appli
45	320	4.8	4724	1	US-08-404-671-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-877-730-7
; Sequence 7, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632e1 Human Phosphatases and Polynucleotides Encoding t
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3453
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-7

Alignment Scores:
Pred. No.:              4.42e-106       Length:        3453
Score:                  1747.50         Matches:       397
Percent Similarity:     54.95%          Conservative:  169
Best Local Similarity:  38.54%          Mismatches:    381
Query Match:            26.46%          Indels:        83
DB:                     4               Gaps:          18

US-09-754-997A-2 (1-1252) x US-09-877-730-7 (1-3453)
```

```
Qy     9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db    43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy    29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db    97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy    48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db   157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy    68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db   208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy    88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db   268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy   108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db   319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy   128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db   361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy   148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db   421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy   168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db   481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy   187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db   541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy   207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db   601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy   227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db   661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCAACAATTATA 714

Qy   247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db   715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy   267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db   775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy   287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db   835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy   306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db   895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy   326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db   955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy   346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db  1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy   366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db  1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy   386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db  1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy   406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db  1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy   426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db  1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy   446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db  1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy   466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db  1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy   486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db  1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy   506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db  1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548

Qy   526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db  1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608

Qy   546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db  1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668

Qy   566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db  1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728

Qy   586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db  1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788

Qy   606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db  1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845

Qy   626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db  1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902

Qy   645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db  1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962

Qy   664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db  1963 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 1989

Qy   684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db  1990 -----------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031

Qy   704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db  2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091

Qy   724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
```

```
Db       2092 GATGGCTATCAGGCA-----------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136

Qy        744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
                |||                  ||||||   ||||||   |||::::::|||::::::
Db       2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196

Qy        760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
              |||:::|||:::|||||||:::|||   |||::::::|||   ||||||      :::|||
Db       2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256

Qy        780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
              :::|||||||||:::|||      |||   ||||||:::|||||||||||||||   |||
Db       2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316

Qy        800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
                    :::|||      |   :::|||:::   ||||||:::|||   ||||||||||||
Db       2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376

Qy        820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
              |||||||||:::   |||                :::   |||:::   ||||||
Db       2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436

Qy        840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
              ||||||||||||:::   |||:::   ||||||      ::::::::::::   :::
Db       2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496

Qy        860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
                 |||      :::   |||   ||||||      |||         :::
Db       2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556

Qy        880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
              |||   |||||||||::::::                         :::|||   :::|||
Db       2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616

Qy        900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
              ||||||   |||   |||   :::      |||   :::            |||
Db       2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676

Qy        920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
              |||:::   |||      |||||||||   ||||||||||||         ::::::
Db       2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736

Qy        940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
                 |||   :::      |||:::|||
Db       2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796

Qy        950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
                                            ||||||   ::::::|||||||||
Db       2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856

Qy        960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
              |||||||||   :::   |||   |||:::|||   |||:::         :::   :::
Db       2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916

Qy        980 SerSerHisArgGluAla-----------------------------------LeuPro 987
              |||      |||::::::                                   ||||||
Db       2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976

Qy        988 GlyLeuSerSerSerGlyThrProGlyAsn 997
                 |||:::|||      ||||||
Db       2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006
```

RESULT 2
US-09-877-730-1
; Sequence 1, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3210
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-1

Alignment Scores:

Pred. No.:	4.66e-106	Length:	3210
Score:	1746.50	Matches:	397
Percent Similarity:	54.95%	Conservative:	169
Best Local Similarity:	38.54%	Mismatches:	381
Query Match:	26.45%	Indels:	83
DB:	4	Gaps:	18

US-09-754-997A-2 (1-1252) x US-09-877-730-1 (1-3210)

```
Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
              |||:::|||:::                  ||||||               |||||||||
Db         43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
                       :::||||||      :::   |||      |||   :::
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
              :::         |||||||||||||||                     |||   :::
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
              |||||||||   |||:::|||      :::   |||::::::   :::   :::|||
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
              |||||||||||||:::::::|||      |||   |||         |||::::::
Db        268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
                                ||||||   |||   |||||||||   :::      |||
Db        319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
                 :::   ||||||   |||   :::   |||:::|||:::         |||   :::
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy        148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
              |||   |||      |||   |||   |||   |||||||||   |||
Db        421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy        168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
                 |||         :::|||||||||||   ::::::   |||:::|||
Db        481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy        187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
              |||:::      ||||||      :::|||||||||   ||||||         |||:::
Db        541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy        207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
              |||:::|||||||||:::|||         |||   :::|||   |||   |||||||||
Db        601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy        227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
              |||||||||               :::   |||:::               |||:::
Db        661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714
```

```
Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774
Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834
Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894
Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954
Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014
Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074
Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128
Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188
Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248
Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308
Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368
Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488
Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1548
Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1608
Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1609 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1668
Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1669 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1728
Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1729 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1788
Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1789 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 1845
Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   1846 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 1902
Qy    645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   1903 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 1962
Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   1963 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 1989
Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   1990 -----------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2031
Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2032 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2091
Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2092 GATGGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2136
Qy    744 IleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2137 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2196
Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2197 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2256
Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db   2257 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2316
Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2317 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2376
Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2377 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2436
Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2437 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2496
Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2497 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2556
Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2557 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2616
Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2617 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2676
Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2677 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2736
Qy    940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db   2737 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2796
Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
```

```
Db     2797 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 2856

Qy      960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979

Db     2857 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 2916

Qy      980 SerSerHisArgGluAla---------------------------------LeuPro 987

Db     2917 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 2976

Qy      988 GlyLeuSerSerSerGlyThrProGlyAsn 997

Db     2977 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3006
```

RESULT 3
US-09-877-730-31
; Sequence 31, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 3874
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-31

Alignment Scores:

Pred. No.:	6.04e-106	Length:	3874
Score:	1746.50	Matches:	397
Percent Similarity:	54.95%	Conservative:	169
Best Local Similarity:	38.54%	Mismatches:	381
Query Match:	26.45%	Indels:	83
DB:	4	Gaps:	18

US-09-754-997A-2 (1-1252) x US-09-877-730-31 (1-3874)

```
Qy        9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28

Db      233 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 286

Qy       29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47

Db      287 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 346

Qy       48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67

Db      347 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 397

Qy       68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87

Db      398 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 457

Qy       88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107

Db      458 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 508

Qy      108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127

Db      509 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 550

Qy      128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147

Db      551 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 610

Qy      148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167

Db      611 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 670

Qy      168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186

Db      671 CACCCTCCTGCAGTCATAACATGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 730

Qy      187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206

Db      731 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 790

Qy      207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226

Db      791 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 850

Qy      227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246

Db      851 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 904

Qy      247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266

Db      905 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 964

Qy      267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286

Db      965 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 1024

Qy      287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305

Db     1025 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 1084

Qy      306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325

Db     1085 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 1144

Qy      326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345

Db     1145 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1204

Qy      346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365

Db     1205 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1264

Qy      366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385

Db     1265 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1318

Qy      386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405

Db     1319 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1378

Qy      406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425

Db     1379 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1438

Qy      426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445

Db     1439 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1498

Qy      446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465

Db     1499 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1558

Qy      466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485

Db     1559 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1618
```

```
Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1619 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1678

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1679 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTT 1738

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1739 CCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCC 1798

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1799 TGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTC 1858

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1859 CGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTAC 1918

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1919 CTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGA 1978

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1979 GTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGC 2035

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeu 644
Db   2036 GTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATT 2092

Qy    645 ValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyr 663
Db   2093 TCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTAC 2152

Qy    664 TrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGly 683
Db   2153 TACAAGGAAGAAGGGCAGCAGGAGAAT--------------------------------- 2179

Qy    684 AspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeu 703
Db   2180 ------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTC 2221

Qy    704 ThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGlu 723
Db   2222 AGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGAC 2281

Qy    724 AspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuPro 743
Db   2282 GATGGCTATCAGGCA----------------GATCAGACTGTCAGCACTCCAGGATGCGTG 2326

Qy    744 IleGlnArgGly---------------ProProLeuProProAlaHisValHisAlaGluSer 759
Db   2327 TCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCT 2386

Qy    760 AsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIle 779
Db   2387 AACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATC 2446

Qy    780 ValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr 799
Db   2447 ATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTAC 2506

Qy    800 TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGlu 819
Db   2507 CTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAA 2566

Qy    820 PheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArg 839
Db   2567 TTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCAT 2626

Qy    840 SerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThr 859
Db   2627 TCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAG 2686

Qy    860 ProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGlu 879
Db   2687 GATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGC 2746

Qy    880 TyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThr 899
Db   2747 TATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACAC 2806

Qy    900 ThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhe 919
Db   2807 CGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATT 2866

Qy    920 PheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspVal 939
Db   2867 GTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTG 2926

Qy    940 IleThrLeuGlnGluThrPheSerAspSer------------------------------ 949
Db   2927 GCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCT 2986

Qy    950 ------------------------------LeuAspValHisAlaValThrGlyIleIle 959
Db   2987 GATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCT 3046

Qy    960 ValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGln 979
Db   3047 GTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGA 3106

Qy    980 SerSerHisArgGluAla------------------------------------LeuPro 987
Db   3107 AGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCT 3166

Qy    988 GlyLeuSerSerSerGlyThrProGlyAsn 997
Db   3167 CGTACCAGTGCCTCCTTAGCTAGTGGAAAT 3196
```

RESULT 4
US-09-877-730-11
; Sequence 11, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2976
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-11

Alignment Scores:

Pred. No.:	1.79e-102	Length:	2976
Score:	1691.50	Matches:	375
Percent Similarity:	55.79%	Conservative:	160
Best Local Similarity:	39.10%	Mismatches:	347
Query Match:	25.61%	Indels:	77
DB:	4	Gaps:	15

US-09-754-997A-2 (1-1252) x US-09-877-730-11 (1-2976)

```
Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107

Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167

Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227

Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287

Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347

Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGC 1407

Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467

Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527

Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587

Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1588 CATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641

Qy    637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
Db   1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701

Qy    656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAsp 674
Db   1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755

Qy    675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694
Db   1756 ---------------------------------------------GGGCCCATTTTCTTG 1770

Qy    695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714
Db   1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830

Qy    715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734
Db   1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875

Qy    735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750
Db   1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935

Qy    751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770
Db   1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995

Qy    771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790
Db   1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055
```

```
Qy    791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810
Db   2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAA 2115

Qy    811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830
Db   2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175

Qy    831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
Db   2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235

Qy    851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
Db   2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295

Qy    871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
Db   2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355

Qy    891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
Db   2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415

Qy    911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
Db   2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475

Qy    931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949
Db   2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535

Qy    950 ---------------------------------------------------------Leu 950
Db   2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595

Qy    951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
Db   2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655

Qy    971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985
Db   2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715

Qy    986 ---------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 5
US-09-877-730-17
; Sequence 17, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 3219
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-17

Alignment Scores:

Pred. No.:	2e-102	Length:	3219
Score:	1691.50	Matches:	375
Percent Similarity:	55.79%	Conservative:	160
Best Local Similarity:	39.10%	Mismatches:	347
Query Match:	25.61%	Indels:	77
DB:	4	Gaps:	15

US-09-754-997A-2 (1-1252) x US-09-877-730-17 (1-3219)

```
Qy     79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db      7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy     99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db     58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy    119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db    100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy    139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db    160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy    159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db    220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGAGTTCAAT 279

Qy    179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db    280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy    198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db    340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy    218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db    400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy    238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db    454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy    258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db    514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy    278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db    574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy    297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db    634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy    317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db    694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy    337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db    754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy    357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db    814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy    377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
```

```
Db    874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927
Qy    397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db    928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987
Qy    417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db    988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047
Qy    437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db   1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107
Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167
Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227
Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287
Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGT 1347
Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1348 CGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGC 1407
Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1408 CAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTG 1467
Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1468 GAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTAC 1527
Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1528 CTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCA 1587
Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1588 CATAGGACGCCCAAAGCT--ACAAGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCAT 1641
Qy    637 ValArg---AlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThr 655
Db   1642 TTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACA 1701
Qy    656 ---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluAlaAsp 674
Db   1702 GCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAAGGGCAGCAGGAGAAT------ 1755
Qy    675 GlyAspArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeu 694
Db   1756 ---------------------------------------------GGGCCCATTTTCTTG 1770
Qy    695 LysLysLysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluVal 714
Db   1771 GATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTG 1830
Qy    715 LysLeuValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThr 734
Db   1831 AGACTCCTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GAT 1875
Qy    735 GluLysAlaProThrProAspLeuProIleGlnArgGly------------ProProLeu 750
Db   1876 CAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCA 1935

Qy    751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLys 770
Db   1936 CCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGG 1995
Qy    771 LysProAspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGly 790
Db   1996 AGGCCTGCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGC 2055
Qy    791 LeuArgAsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGly 810
Db   2056 CTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAA 2115
Qy    811 GlyLeuLysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAsp 830
Db   2116 GGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCC 2175
Qy    831 GlyProPheGlySerValValGluArgSerThrLeuProAspArgProSerThrProPro 850
Db   2176 AGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCA 2235
Qy    851 SerAspLeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProPro 870
Db   2236 GTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCT 2295
Qy    871 ThrGluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGln 890
Db   2296 GATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGG 2355
Qy    891 ProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHis 910
Db   2356 ATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAA 2415
Qy    911 GlyLeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyPro 930
Db   2416 AACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAA 2475
Qy    931 GlyProPheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--- 949
Db   2476 GGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAAT 2535
Qy    950 -----------------------------------------------------------Leu 950
Db   2536 CAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTG 2595
Qy    951 AspValHisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeu 970
Db   2596 GACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTC 2655
Qy    971 AlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------- 985
Db   2656 ATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACG 2715
Qy    986 --------------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2716 GCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2772
```

RESULT 6
US-09-877-730-5
; Sequence 5, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08

```
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 2715
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-5

Alignment Scores:
Pred. No.:              3.16e-82     Length:         2715
Score:                  1384.00      Matches:        334
Percent Similarity:     46.11%       Conservative:   140
Best Local Similarity:  32.49%       Mismatches:     310
Query Match:            20.96%       Indels:         244
DB:                     4            Gaps:           14

US-09-754-997A-2 (1-1252) x US-09-877-730-5 (1-2715)

Qy      9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db     43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96

Qy     29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db     97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156

Qy     48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db    157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207

Qy     68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db    208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267

Qy     88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db    268 AACGGCTCTTTATACATCAGT---------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318

Qy    108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db    319 -----------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAATATGGA 360

Qy    128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db    361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420

Qy    148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db    421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy    168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db    481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy    187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db    541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy    207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db    601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy    227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
```

```
Db       1554 ------------------------------------------------------------ 1554
Qy        646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db       1554 ------------------------------------------------------------ 1554
Qy        666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db       1554 ------------------------------------------------------------ 1554
Qy        686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db       1554 ------------------------------------------------------------ 1554
Qy        706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db       1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602
Qy        726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db       1603 TATCAGGCA-----------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647
Qy        746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db       1648 CGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACC 1707
Qy        762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db       1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767
Qy        782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db       1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827
Qy        802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db       1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887
Qy        822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db       1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947
Qy        842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db       1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007
Qy        862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
Db       2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067
Qy        882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db       2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127
Qy        902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
Db       2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187
Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
Db       2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247
Qy        942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
Db       2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307
Qy        950 ------------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
Db       2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367
Qy        962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
Db       2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427

Qy        982 HisArgGluAla----------------------------------LeuProGlyLeu 989
Db       2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487
Qy        990 SerSerSerGlyThrProGlyAsn 997
Db       2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

RESULT 7
US-09-877-730-9
; Sequence 9, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 2958
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-9

Alignment Scores:

Pred. No.:	3.56e-82	Length:	2958
Score:	1384.00	Matches:	334
Percent Similarity:	46.11%	Conservative:	140
Best Local Similarity:	32.49%	Mismatches:	310
Query Match:	20.96%	Indels:	244
DB:	4	Gaps:	14

US-09-754-997A-2 (1-1252) x US-09-877-730-9 (1-2958)

```
Qy          9 GlyLeuLeuValLeuThrPheCysLeuLeuSerAlaArgGlyGluLeuProLeuProGln 28
Db         43 GGGATGCTGCTCCGCGCGCTCCTGCTCCTGCTGMTGCTCAGT------CCTTTGCCAGGA 96
Qy         29 GluThrThrVal---LysLeuSerCysAspGluGlyProLeuGlnValIleLeuGlyPro 47
Db         97 GTGTGGTGCTTTAGCGAACTGTCTTTTGTAAAAGAACCACAGGATGTAACTGTCACAAGA 156
Qy         48 GluGlnAlaValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArg 67
Db        157 AAGGACCCAGTCGTTTTAGATTGCCAGGCTCACGGAGAAGTT---------CCTATTAAG 207
Qy         68 ValThrTrpSerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeuLeuPro 87
Db        208 GTCACATGGTTGAAAAATGGAGCAAAAATGTCTGAAAATAAACGGATCGAGGTTCTTTCT 267
Qy         88 AsnGlySerLeuTrpLeuSerSerProLeuGluGlnGluAspSerAspAspGluGluAla 107
Db        268 AACGGCTCTTTATACATCAGT----------GAGGTGGAAGGCAGGCGAGGAGAGCAGTCC 318
Qy        108 LeuArgIleTrpLysValThrGluGlySerTyrSerCysLeuAlaHisSerProLeuGly 127
Db        319 ------------------GATGAAGGATTTTATCAGTGCTTGGCAATGAACAAAATATGGA 360
Qy        128 ValValAlaSerGlnValAlaValValLysLeuAlaThrLeuGluAspPheSerLeuHis 147
Db        361 GCCATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTTCTGCATTTGAAGTCCAG 420
```

```
Qy    148 ProGluSerGlnIleValGluGluAsnGlyThrAlaArgPheGluCysHisThrLysGly 167
Db    421 CCAATTTCCACTGAGGTCCACGAAGGTGGAGTTGCTCGATTTGCATGCAAGATTTCATCC 480

Qy    168 LeuProAlaProIleIleThrTrpGluLysAspGlnValThrValPro---GluGluPro 186
Db    481 CACCCTCCTGCAGTCATAACATGGGAGTTCAATCGGACAACTCTACCTATGACTATGGAC 540

Qy    187 ArgLeuIleThrLeuProLysTrpLeuLeuGlnIleLeuAspValGlnAspSerAspAla 206
Db    541 AGGATAACTGCCCTACCAACAGGAGTATTGCAGATCTATGATGTCAGCCAAAGGGATTCT 600

Qy    207 GlySerTyrArgCysValAlaThrAsnSerAlaArgGlnArgPheSerGlnGluAlaSer 226
Db    601 GGAAATTATCGTTGTATTGCTGCCACTGTAGCCCACCGACGTAAAAGTATGGAGGCCTCG 660

Qy    227 LeuThrValAlaLeuArgGlySerLeuGluAlaThrArgGlyGlnAspValValIleVal 246
Db    661 CTAACTGTGATT------CCAGCTAAGGAGTCAAAATCCTTCCACACACCARCAATTATA 714

Qy    247 AlaAlaProGluAsnThrThrValValSerGlyGlnAsnValValMetGluCysValAla 266
Db    715 GCAGGTCCACAGAACATAACAACATCTCTTCATCAGACTGTAGTTTTGGAATGCATGGCC 774

Qy    267 SerAlaAspProThrProPheValSerTrpValArgGlnAspGlyLysProIleSerThr 286
Db    775 ACAGGAAATCCCAAACCAATCATTTCTTGGAGCCGCCTTGATCACAAATCCATTGATGTC 834

Qy    287 ---AspValIleValLeuGlyArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHis 305
Db    835 TTTAATACTCGGGTACTTGGAAATGGTAATCTCATGATATCTGATGTCAGGCTACAACAT 894

Qy    306 SerGlyValTyrValCysArgAlaAsnLysProLeuThrArgAspPheAlaThrAlaAla 325
Db    895 GCTGGAGTATATGTTTGTCGGGCCACTACCCCTGGCACACGCAACTTTACAGTTGCTATG 954

Qy    326 AlaGluLeuArgValLeuAlaAlaProAlaIleSerGlnAlaProGluAlaLeuSerArg 345
Db    955 GCAACTTTAACTGTATTAGCTCCTCCTTCATTTGTTGAATGGCCAGAAAGTTTAACAAGG 1014

Qy    346 ThrArgAlaSerThrAlaArgPheValCysArgAlaSerGlyGluProArgProAlaLeu 365
Db   1015 CCTCGAGCTGGCACTGCTCGATTTGTGTGTCAGGCAGAAGGAATCCCCTCTCCCAAGATG 1074

Qy    366 HisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGlyGly 385
Db   1075 TCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC------ 1128

Qy    386 GlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysVal 405
Db   1129 AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGCATG 1188

Qy    406 AlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArgGlu 425
Db   1189 GCTGAGAATAGCCAAGGATCTATTTATCTAGAGCCAGACTGACTGTAGTGATGTCAGAA 1248

Qy    426 GlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeu 445
Db   1249 GACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATTCTT 1308

Qy    446 ValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyr 465
Db   1309 TTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACACTAC 1368

Qy    466 GlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThr 485
Db   1369 ATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACT 1428

Qy    486 GluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyr 505
Db   1429 CATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATAT 1488

Qy    506 SerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspVal 525
Db   1489 ATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGAC 1548

Qy    526 ProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAla 545
Db   1549 CCCAGA------------------------------------------------------ 1554

Qy    546 TrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyr 565
Db   1554 ------------------------------------------------------------ 1554

Qy    566 GlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeu 585
Db   1554 ------------------------------------------------------------ 1554

Qy    586 ThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGly 605
Db   1554 ------------------------------------------------------------ 1554

Qy    606 AlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGln 625
Db   1554 ------------------------------------------------------------ 1554

Qy    626 SerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuVal 645
Db   1554 ------------------------------------------------------------ 1554

Qy    646 ValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   1554 ------------------------------------------------------------ 1554

Qy    666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   1554 ------------------------------------------------------------ 1554

Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   1554 ------------------------------------------------------------ 1554

Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   1555 ------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGC 1602

Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   1603 TATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTT 1647

Qy    746 ArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsnSer 761
Db   1648 CGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACC 1707

Qy    762 SerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsn 781
Db   1708 TCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAAC 1767

Qy    782 TyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThr 801
Db   1768 TACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAA 1827

Qy    802 SerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAla 821
Db   1828 ACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCC 1887

Qy    822 ValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThr 841
Db   1888 GTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACT 1947

Qy    842 LeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSer 861
Db   1948 CTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGAC 2007

Qy    862 ThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeu 881
```

```
Db       2008 ACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACT 2067

Qy        882 IleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGlu 901
Db       2068 ATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAA 2127

Qy        902 GlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLys 921
Db       2128 GGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAG 2187

Qy        922 MetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThr 941
Db       2188 ATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTA 2247

Qy        942 LeuGlnGluThrPheSerAspSer------------------------------------ 949
Db       2248 CTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCC 2307

Qy        950 ------------------------LeuAspValHisAlaValThrGlyIleIleValGly 961
Db       2308 AAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGT 2367

Qy        962 ValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSer 981
Db       2368 GTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAA 2427

Qy        982 HisArgGluAla---------------------------------LeuProGlyLeu 989
Db       2428 GCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACC 2487

Qy        990 SerSerSerGlyThrProGlyAsn 997
Db       2488 AGTGCCTCCTTAGCTAGTGGAAAT 2511
```

RESULT 8
US-09-877-730-15
; Sequence 15, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-15

Alignment Scores:

Pred. No.:	1.2e-78	Length:	2481
Score:	1329.00	Matches:	312
Percent Similarity:	46.29%	Conservative:	131
Best Local Similarity:	32.60%	Mismatches:	276
Query Match:	20.12%	Indels:	238
DB:	4	Gaps:	11

US-09-754-997A-2 (1-1252) x US-09-877-730-15 (1-2481)

```
Qy         79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db          7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy         99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db         58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy        119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db        100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy        139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db        160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy        159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db        220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGGAGTTCAAT 279

Qy        179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db        280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy        198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db        340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy        218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db        400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy        238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db        454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy        258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db        514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy        278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db        574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy        297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db        634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy        317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db        694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy        337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db        754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy        357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db        814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy        377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db        874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy        397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db        928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy        417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db        988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy        437 ThrProLeuSerSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db       1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107
```

```
Qy    457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db   1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167
Qy    477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db   1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227
Qy    497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db   1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287
Qy    517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db   1288 GTGACACAGAATACTCTAGAGGATGACCCCAGA--------------------------- 1320
Qy    537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db   1320 ------------------------------------------------------------ 1320
Qy    557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db   1320 ------------------------------------------------------------ 1320
Qy    577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db   1320 ------------------------------------------------------------ 1320
Qy    597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db   1320 ------------------------------------------------------------ 1320
Qy    617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db   1320 ------------------------------------------------------------ 1320
Qy    637 ValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGln 656
Db   1320 ------------------------------------------------------------ 1320
Qy    657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
Db   1320 ------------------------------------------------------------ 1320
Qy    677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
Db   1320 ------------------------------------------------------------ 1320
Qy    697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
Db   1321 ---------------------------------------AGAAAATATCATGTGAGACTC 1341
Qy    717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
Db   1342 CTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GATCAGACT 1386
Qy    737 AlaProThrProAspLeuProIleGlnArgGly------------ProProLeuProPro 752
Db   1387 GTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCC 1446
Qy    753 AlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysPro 772
Db   1447 CACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCT 1506
Qy    773 AspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArg 792
Db   1507 GCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAG 1566
Qy    793 AsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeu 812
Db   1567 AATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTA 1626
Qy    813 LysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyPro 832
Db   1627 GAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCT 1686
Qy    833 PheGlySerValValGluArgSerThrLeuProAspArgProSerThrProProSerAsp 852
Db   1687 TGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGA 1746
Qy    853 LeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGlu 872
Db   1747 GTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGC 1806
Qy    873 ProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGlu 892
Db   1807 CCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCA 1866
Qy    893 HisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeu 912
Db   1867 GGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTG 1926
Qy    913 GluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyPro 932
Db   1927 GTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCC 1986
Qy    933 PheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--------- 949
Db   1987 TTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGG 2046
Qy    950 ---------------------------------------------------LeuAspVal 952
Db   2047 CCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAA 2106
Qy    953 HisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCys 972
Db   2107 AAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGT 2166
Qy    973 MetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------------- 985
Db   2167 GTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAG 2226
Qy    986 ---------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db   2227 AATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2277
```

RESULT 9
US-09-877-730-19
; Sequence 19, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 2724
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-19

Alignment Scores:

Pred. No.:	1.36e-78	Length:	2724
Score:	1329.00	Matches:	312
Percent Similarity:	46.29%	Conservative:	131

```
Best Local Similarity:  32.60%      Mismatches:   276
Query Match:            20.12%      Indels:       238
DB:                     4           Gaps:         11

US-09-754-997A-2 (1-1252) x US-09-877-730-19 (1-2724)

Qy      79 GluHisGluAsnLeuHisLeuLeuProAsnGlySerLeuTrpLeuSerSerProLeuGlu 98
Db       7 GAAAATAAACGGATCGAGGTTCTTTCTAACGGCTCTTTATACATCAGT---------GAG 57

Qy      99 GlnGluAspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyr 118
Db      58 GTGGAAGGCAGGCGAGGAGAGCAGTCC------------------GATGAAGGATTTTAT 99

Qy     119 SerCysLeuAlaHisSerProLeuGlyValValAlaSerGlnValAlaValValLysLeu 138
Db     100 CAGTGCTTGGCAATGAACAAATATGGAGCCATTCTTAGTCAAAAAGCTCATCTTGCCTTA 159

Qy     139 AlaThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThr 158
Db     160 TCAACTATTTCTGCATTTGAAGTCCAGCCAATTTCCACTGAGGTCCACGAAGGTGGAGTT 219

Qy     159 AlaArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAsp 178
Db     220 GCTCGATTTGCATGCAAGATTTCATCCCACCCTCCTGCAGTCATAACATGGAGTTCAAT 279

Qy     179 GlnValThrValPro---GluGluProArgLeuIleThrLeuProLysTrpLeuLeuGln 197
Db     280 CGGACAACTCTACCTATGACTATGGACAGGATAACTGCCCTACCAACAGGAGTATTGCAG 339

Qy     198 IleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSerAla 217
Db     340 ATCTATGATGTCAGCCAAAGGGATTCTGGAAATTATCGTTGTATTGCTGCCACTGTAGCC 399

Qy     218 ArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGluAla 237
Db     400 CACCGACGTAAAAGTATGGAGGCCTCGCTAACTGTGATT------CCAGCTAAGGAGTCA 453

Qy     238 ThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSerGly 257
Db     454 AAATCCTTCCACACACCARCAATTATAGCAGGTCCACAGAACATAACAACATCTCTTCAT 513

Qy     258 GlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrpVal 277
Db     514 CAGACTGTAGTTTTGGAATGCATGGCCACAGGAAATCCCAAACCAATCATTTCTTGGAGC 573

Qy     278 ArgGlnAspGlyLysProIleSerThr---AspValIleValLeuGlyArgThrAsnLeu 296
Db     574 CGCCTTGATCACAAATCCATTGATGTCTTTAATACTCGGGTACTTGGAAATGGTAATCTC 633

Qy     297 LeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArgAlaAsnLysPro 316
Db     634 ATGATATCTGATGTCAGGCTACAACATGCTGGAGTATATGTTTGTCGGGCCACTACCCCT 693

Qy     317 LeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAlaAlaProAlaIle 336
Db     694 GGCACACGCAACTTTACAGTTGCTATGGCAACTTTAACTGTATTAGCTCCTCCTTCATTT 753

Qy     337 SerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArgPheValCysArg 356
Db     754 GTTGAATGGCCAGAAAGTTTAACAAGGCCTCGAGCTGGCACTGCTCGATTTGTGTGTCAG 813

Qy     357 AlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIleProLeuArgPro 376
Db     814 GCAGAAGGAATCCCCTCTCCCAAGATGTCATGGTTGAAAAATGGAAGGAAGATACATTCG 873

Qy     377 AsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThrGlnIleGlyLeu 396
Db     874 AATGGTAGAATTAAAATGTAC------AACAGTAAATTGGTAATTAACCAGATTATTCCT 927

Qy     397 GlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThrAlaCysAlaAla 416
Db     928 GAAGATGATGCTATTTATCAGTGCATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGA 987

Qy     417 AlaProLeuAlaValValValArgGluGlyLeuProSerAlaProThrArgValThrAla 436
Db     988 GCCAGACTGACTGTAGTGATGTCAGAAGACAGACCCAGTGCTCCCTATAATGTACATGCT 1047

Qy     437 ThrProLeuSerSerSerValLeuValAlaTrpGluArgProGluLeuHisSerGlu 456
Db    1048 GAAACCATGTCAAGCTCAGCCATTCTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGAC 1107

Qy     457 GlnIleIleGlyPheSerLeuHisTyrGlnLysAlaArgGlyValAspAsnValGluTyr 476
Db    1108 AAAGTCATTGCCTATTCTGTACACTACATGAAAGCAGAAGGTTTAAATAATGAAGAGTAT 1167

Qy     477 GlnPheAlaValAsnAsnAspThrThrGluLeuGlnValArgAspLeuGluProAsnThr 496
Db    1168 CAAGTAGTCATCGGAAATGACACAACTCATTATATTATTGATGACTTAGAGCCTGCCAGC 1227

Qy     497 AspTyrGluPheTyrValValAlaTyrSerGlnLeuGlyAlaSerArgThrSerSerPro 516
Db    1228 AATTATACTTTCTACATTGTAGCATATATGCCAATGGGAGCCAGCCAGATGTCTGACCAT 1287

Qy     517 AlaLeuValHisThrLeuAspAspValProSerAlaAlaProGlnLeuThrLeuSerSer 536
Db    1288 GTGACACAGAATACTCTAGAGGATGACCCCAGA--------------------------- 1320

Qy     537 ProAsnProSerAspIleArgValAlaTrpLeuProLeuProSerSerLeuSerAsnGly 556
Db    1320 ------------------------------------------------------------ 1320

Qy     557 GlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLysGluAspGlnValPheSerThr 576
Db    1320 ------------------------------------------------------------ 1320

Qy     577 GluValProGlyAsnGluThrGlnLeuThrLeuAsnSerLeuGlnProAsnLysValTyr 596
Db    1320 ------------------------------------------------------------ 1320

Qy     597 ArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGlyValProSerGlnTrpMetGln 616
Db    1320 ------------------------------------------------------------ 1320

Qy     617 HisArgThrProGlyValHisAsnGlnSerHisValProPheAlaProAlaGluLeuLys 636
Db    1320 ------------------------------------------------------------ 1320

Qy     637 ValArgAlaLysMetGluSerLeuValValSerTrpGlnProProProHisProThrGln 656
Db    1320 ------------------------------------------------------------ 1320

Qy     657 IleSerGlyTyrLysLeuTyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAsp 676
Db    1320 ------------------------------------------------------------ 1320

Qy     677 ArgProProGlyGlyArgGlyAspGlnAlaTrpAspValGlyProValArgLeuLysLys 696
Db    1320 ------------------------------------------------------------ 1320

Qy     697 LysValLysGlnTyrGluLeuThrGlnLeuValProGlyArgProTyrGluValLysLeu 716
Db    1321 ------------------------------------AGAAAATATCATGTGAGACTC 1341

Qy     717 ValAlaPheAsnLysHisGluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLys 736
Db    1342 CTGGCTTACAACAACATAGACGATGGCTATCAGGCA---------------GATCAGACT 1386

Qy     737 AlaProThrProAspLeuProIleGlnArgGly------------ProProLeuProPro 752
Db    1387 GTCAGCACTCCAGGATGCGTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCC 1446

Qy     753 AlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrpLysLysPro 772
Db    1447 CACCATCTCTATGCGAAGGCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCT 1506
```

```
Qy       773 AspPheThrThrValLysIleValAsnTyrThrValArgPheGlyProTrpGlyLeuArg 792
Db      1507 GCATTCACCGCTGCACAAATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAG 1566
Qy       793 AsnAlaSerLeuValThrTyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeu 812
Db      1567 AATGCTTCTTTGGTTCTGTACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTA 1626
Qy       813 LysProPheThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyPro 832
Db      1627 GAACCAAACACCAAATACGAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCT 1686
Qy       833 PheGlySerValValGluArgSerThrLeuProAspArgProSerThrProProSerAsp 852
Db      1687 TGGAGCCCTGTAGTCTACCATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGA 1746
Qy       853 LeuArgLeuSerProLeuThrProSerThrValArgLeuHisTrpCysProProThrGlu 872
Db      1747 GTAAAAGTGACATTAATAGAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGC 1806
Qy       873 ProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGlu 892
Db      1807 CCAGAAACAGTTGTGACCCGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCA 1866
Qy       893 HisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeu 912
Db      1867 GGAGAGTGGCAGGTCTTACACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTG 1926
Qy       913 GluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGlyPro 932
Db      1927 GTAGCAGGAAATGTGTACATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCC 1986
Qy       933 PheSerArgLeuGlnAspValIleThrLeuGlnGluThrPheSerAspSer--------- 949
Db      1987 TTTTCAAATTCTGTGGAGCTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGG 2046
Qy       950 -------------------------------------------------LeuAspVal 952
Db      2047 CCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAA 2106
Qy       953 HisAlaValThrGlyIleIleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCys 972
Db      2107 AAATCAATGACTGGCATTGCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGT 2166
Qy       973 MetCysAlaGlyLeuArgGlnSerSerHisArgGluAla--------------------- 985
Db      2167 GTTCTCATCTTGATATACCGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAG 2226
Qy       986 ---------------LeuProGlyLeuSerSerSerGlyThrProGlyAsn 997
Db      2227 AATGGAACTCAACAGTTACCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 2277
```

RESULT 10
US-09-877-730-21
; Sequence 21, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-21

```
Alignment Scores:
Pred. No.:               1.59e-66      Length:         2139
Score:                   1144.00       Matches:        252
Percent Similarity:      54.69%        Conservative:   115
Best Local Similarity:   37.56%        Mismatches:     240
Query Match:             17.32%        Indels:         64
DB:                      4             Gaps:           10
```

US-09-754-997A-2 (1-1252) x US-09-877-730-21 (1-2139)

```
Qy       365 LeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGly 384
Db         1 ATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC--- 57
Qy       385 GlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCys 404
Db        58 ---AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGC 114
Qy       405 ValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArg 424
Db       115 ATGGCTGAGAATAGCCAAGGATCTATTTTATCTAGAGCCAGACTGACTGTAGTGATGTCA 174
Qy       425 GluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerVal 444
Db       175 GAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATT 234
Qy       445 LeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHis 464
Db       235 CTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACAC 294
Qy       465 TyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThr 484
Db       295 TACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACA 354
Qy       485 ThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAla 504
Db       355 ACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCA 414
Qy       505 TyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAsp 524
Db       415 TATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGAT 474
Qy       525 ValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgVal 544
Db       475 GTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATC 534
Qy       545 AlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGlu 564
Db       535 TCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCT 594
Qy       565 TyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGln 584
Db       595 TTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAG 654
Qy       585 LeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThr 604
Db       655 TACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACC 714
Qy       605 GlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsn 624
Db       715 AGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACA 771
Qy       625 GlnSerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSer 643
Db       772 AGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACC 828
Qy       644 LeuValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeu 662
```

```
Db        829 ATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTG 888

Qy        663 TyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArg 682
              |||:::  |||   |||  :::|||
Db        889 TACTACAAGGAAGAAGGGCAGCAGGAGAAT------------------------------ 918

Qy        683 GlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGlu 702
                                   ||||||:::  |||    |||        |||
Db        919 ---------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACT 957

Qy        703 LeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHis 722
              |||:::  |||  |||  |||   |||  |||:::|||:::|||:::|||
Db        958 CTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATA 1017

Qy        723 GluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeu 742
              :::|||||||||  |||                 :::     ||||||
Db       1018 GACGATGGCTATCAGGCA--------------GATCAGACTGTCAGCACTCCAGGATGC 1062

Qy        743 ProIleGlnArgGly------------ProProLeuProProAlaHisValHisAlaGlu 758
                        |||           ||||||  ||||||   |||::::::|||:::
Db       1063 GTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAG 1122

Qy        759 SerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLys 778
              :::|||:::|||:::||||||:::|||   |||::::::|||   ||||||     :::
Db       1123 GCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAA 1182

Qy        779 IleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThr 798
              |||:::|||||||||:::|||       |||  ||||||:::||||||||||||||
Db       1183 ATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTG 1242

Qy        799 TyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyr 818
              |||     :::|||        :::|||:::   ||||||:::|||  |||||||||
Db       1243 TACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATAC 1302

Qy        819 GluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGlu 838
              ||||||||||||:::   |||       :::    |||:::      ||||||
Db       1303 GAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTAC 1362

Qy        839 ArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeu 858
                 ||||||||||||:::  |||:::   ||||||      ::::::::::::   :::
Db       1363 CATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATA 1422

Qy        859 ThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleVal 878
                       |||      :::  |||  ||||||     |||        :::
Db       1423 GAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACC 1482

Qy        879 GluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeu 898
                 |||   |||||||||:::::::                  :::|||   :::|||
Db       1483 CGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTA 1542

Qy        899 ThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyr 918
                    ||||||   |||   |||  :::      |||   :::              |||
Db       1543 CACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTAC 1602

Qy        919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
                 |||:::  |||      |||||||||   ||||||||||||            :::
Db       1603 ATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAG 1662

Qy        939 ValIleThrLeuGlnGluThrPheSerAspSer--------------------------- 949
              :::      |||   :::      |||:::|||
Db       1663 CTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCT 1722

Qy        950 ---------------------------------LeuAspValHisAlaValThrGlyIle 958
                                               ||||||      ::::::|||||||||
Db       1723 GCTGATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATT 1782

Qy        959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
              ||||||||||   :::  |||  |||:::|||  |||:::           :::
Db       1783 GCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATAC 1842

Qy        979 GlnSerSerHisArgGluAla-----------------------------------Leu 986
              :::|||       |||::::::                                   |||
Db       1843 CGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTA 1902

Qy        987 ProGlyLeuSerSerSerGlyThrProGlyAsn 997
              |||          |||:::|||         ||||||
Db       1903 CCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 1935
```

```
RESULT 11
US-09-877-730-27
; Sequence 27, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 2382
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-27

Alignment Scores:
Pred. No.:               1.85e-66       Length:         2382
Score:                   1144.00        Matches:        252
Percent Similarity:      54.69%         Conservative:   115
Best Local Similarity:   37.56%         Mismatches:     240
Query Match:             17.32%         Indels:         64
DB:                      4              Gaps:           10

US-09-754-997A-2 (1-1252) x US-09-877-730-27 (1-2382)

Qy        365 LeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGly 384
              :::   ||||||   :::|||       :::      ||||||||:::|||:::
Db          1 ATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC--- 57

Qy        385 GlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCys 404
                          |||||||||  ||||||      :::|||          |||||||||
Db         58 ---AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGC 114

Qy        405 ValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArg 424
              :::||||||||||||   |||:::      :::   |||   |||   ||||||:::
Db        115 ATGGCTGAGAATAGCCAAGGATCTATTTATCTAGAGCCAGACTGACTGTAGTGATGTCA 174

Qy        425 GluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerVal 444
              |||         ||||||||||||         |||   |||      :::|||||||||::::::
Db        175 GAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATT 234

Qy        445 LeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHis 464
              |||:::||||||||||||||||       :::|||::::::|||   :::|||:::|||
Db        235 CTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACAC 294

Qy        465 TyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThr 484
              |||   ||||||   |||::::::|||   |||||||||      :::   |||||||||
Db        295 TACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACA 354

Qy        485 ThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAla 504
              |||      :::   ||||||||||||   ::::::|||   ||||||:::||||||
Db        355 ACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCA 414

Qy        505 TyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAsp 524
              |||      :::||||||||||||:::   |||            :::||||||:::|||
Db        415 TATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGAT 474
```

```
Qy      525 ValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgVal 544
            ||||||         |||:|:::::::|||:::|||    :::|||:::||||||    :::
Db      475 GTTCCCCTGAGACCTCCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATC 534

Qy      545 AlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGlu 564
            :::|||||||||:::|||:|:             |||||||||:::    |||::::::
Db      535 TCCTGGCTGCCAATCCCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCT 594

Qy      565 TyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGln 584
            :::   |||      |||:|:    :::          |||:::||||||         :::
Db      595 TTCCGCCTAAGTACTGAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAG 654

Qy      585 LeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThr 604
               |||      |||:|:|||:::      ||||||    |||||||||:::|||    |||
Db      655 TACCTTTTGGAAGGCCTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACC 714

Qy      605 GlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsn 624
               |||   |||       |||   |||      ||||||||||||
Db      715 AGAGTGGGGCTGGGAGAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACA 771

Qy      625 GlnSerHisValProPheAlaProAlaGluLeuLysValArg---AlaLysMetGluSer 643
                           |||   :::|||   ||||||   :::                  :::
Db      772 AGCGTGAAAGCCCCTAAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACC 828

Qy      644 LeuValValSerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeu 662
            :::   |||   ||||||                |||       |||   ||||||||||||
Db      829 ATTTCTGTGAGGTGGCAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTG 888

Qy      663 TyrTrpGlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArg 682
            |||:::   |||   |||   :::|||
Db      889 TACTACAAGGAAGAAGGGCAGCAGGAGAAT------------------------------ 918

Qy      683 GlyAspGlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGlu 702
                                 |||||||:::      |||          |||
Db      919 ---------------------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACT 957

Qy      703 LeuThrGlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHis 722
            |||:::   |||   |||   |||   |||   |||:::|||:::|||:::|||
Db      958 CTCAGTGGCTTAGACCCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATA 1017

Qy      723 GluAspGlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeu 742
            :::||||||||||   |||                  :::         ||||||
Db     1018 GACGATGGCTATCAGGCA-----------------GATCAGACTGTCAGCACTCCAGGATGC 1062

Qy      743 ProIleGlnArgGly---------------ProProLeuProProAlaHisValHisAlaGlu 758
                     |||            ||||||   ||||||   |||::::::|||:::
Db     1063 GTGTCTGTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAG 1122

Qy      759 SerAsnSerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLys 778
            :::|||:::|||:::||||||:::|||   |||::::::|||   ||||||      :::
Db     1123 GCTAACACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAA 1182

Qy      779 IleValAsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThr 798
            |||:::||||||||||:::|||      |||   ||||||:::||||||||||||||||
Db     1183 ATCATTAACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTG 1242

Qy      799 TyrTyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyr 818
            |||      :::|||         :::|||:::   |||||||:::|||   |||||||||
Db     1243 TACCTTCAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATAC 1302

Qy      819 GluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGlu 838
            |||||||||||:::   |||            :::      |||:::      ||||||
Db     1303 GAATTTGCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTAC 1362

Qy      839 ArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeu 858
               |||||||||||:::   |||:::   ||||||      ::::::::::::      :::
Db     1363 CATTCTACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATA 1422

Qy      859 ThrProSerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleVal 878
                     |||      :::   |||   ||||||      |||         :::
Db     1423 GAGGATGACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACC 1482

Qy      879 GluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeu 898
                |||   ||||||||||::::::                        :::|||   :::|||
Db     1483 CGCTATACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTA 1542

Qy      899 ThrThrGluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyr 918
                  ||||||   |||      |||   :::      |||   :::         |||
Db     1543 CACCGTGAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTAC 1602

Qy      919 PhePheLysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAsp 938
                  |||:::   |||      |||||||||   ||||||||||||         :::
Db     1603 ATTGTCAAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAG 1662

Qy      939 ValIleThrLeuGlnGluThrPheSerAspSer------------------------- 949
            :::      |||   :::      |||:::|||
Db     1663 CTGGCAGTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCT 1722

Qy      950 ------------------------------LeuAspValHisAlaValThrGlyIle 958
                                          ||||||      ::::::|||||||||
Db     1723 GCTGATGCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATT 1782

Qy      959 IleValGlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArg 978
                |||||||||   :::   |||   |||:::|||   |||:::         :::
Db     1783 GCTGTAGGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATAC 1842

Qy      979 GlnSerSerHisArgGluAla--------------------------------------Leu 986
            :::|||      |||::::::                                    |||
Db     1843 CGAAGTAAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTA 1902

Qy      987 ProGlyLeuSerSerSerGlyThrProGlyAsn 997
            |||      |||:::|||         ||||||
Db     1903 CCTCGTACCAGTGCCTCCTTAGCTAGTGGAAAT 1935
```

```
RESULT 12
US-09-041-886-24
; Sequence 24, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4608 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;       LOCATION:  1..4342
US-09-041-886-24

Alignment Scores:
Pred. No.:              8.57e-53      Length:         4608
Score:                  943.00        Matches:        357
Percent Similarity:     36.29%        Conservative:   171
Best Local Similarity:  24.54%        Mismatches:     547
Query Match:            14.28%        Indels:         381
DB:                     3             Gaps:           47

US-09-754-997A-2 (1-1252) x US-09-041-886-24 (1-4608)

Qy     51 ValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArgValThrTrp 70
Db    169 GTCCTCCTCGACTGCTCC---GCGGAGTCCGACCGAGGAGTTCCAGTG---ATCAAGTGG 222

Qy     71 SerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeu--------------- 85
Db    223 AAGAAAGATGGC------------------------ATTCATCTGGCCTTGGGAATGGAT 258

Qy     86 ---------------LeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
Db    259 GAAAGGAAGCAGCAACTTTCAAATGGGTCTCTGCTGATACAAAACATACTTCATTCC--- 315

Qy    101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
Db    316 ------------------------AGACACCACAAGCCAGATGAGGGACTTTACCAATGT 351

Qy    121 LeuAlaHis---SerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAla 139
Db    352 GAGGCATCTTTAGGAGATTCTGGCTCAATTATTAGTCGGACAGCAAAAGTTGCAGTAGCA 411

Qy    140 ThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAla 159
Db    412 GGACCACTGAGGTTCCTTTCACAGACAGAATCTGTCACAGCCTTCATGGGAGACACAGTG 471

Qy    160 ArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGln 179
Db    472 CTACTCAAGTGTGAAGTCATTGGGGAGCCCATGCCAACAATCCACTGGCAGAAGAACCAA 531

Qy    180 ---------ValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeu 196
Db    532 CAAGACCTGACTCCAATCCCAGGTGACTCCCGAGTGGTGGTCTTGCCCTCTGGAGCATTG 591

Qy    197 GlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSer 216
Db    592 CAGATCAGCCGACTCCAACCGGGGGACATTGGAATTTACCGATGCTCAGCTCGAAATCCA 651

Qy    217 AlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGlu 236
Db    652 GCCAGCTCAAGAACAGGAAATGAAGCAGAAGTCAGAATTTTATCAGATCCAGGACTGCAT 711

Qy    237 AlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSer 256
Db    712 ------------AGACAGCTGTATTTTCTGCAAAGACCATCCAATGTAGTAGCCATTGAA 759

Qy    257 GlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrp 276
Db    760 GGAAAAGATGCTGTCCTGGAATGTTGTGTTTCTGGCTATCCTCCACCAAGTTTTACCTGG 819

Qy    277 ValArgGlnAspGlyLysProIle------------SerThrAspValIleValLeuGly 292
Db    820 TTACGA------GGCGAGGAAGTCATCCAACTCAGGTCTAAAAAGTATTCTTTATTGGGT 873

Qy    293 ArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArg 312
Db    874 GGAAGCAACTTGCTTATCTCCAATGTGACAGATGATGACAGTGGAATGTATACCTGTGTT 933

Qy    313 AlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaAlaGluLeuArgValLeuAla 332
Db    934 GTCACA------TATAAAAATGAGAATATTAGTGCCTCTGCAGAGCTCACAGTCTTGGTT 987

Qy    333 AlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArg 352
Db    988 CCGCCATGGTTTTTAAATCATCCTTCCAACCTGTATGCCTATGAAAGCATGGATATTGAG 1047

Qy    353 PheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIle 372
Db   1048 TTTGAATGTACAGTCTCTGGAAAGCCTGTGCCCACTGTGAATTGGATGAAGAATGGAGAT 1107

Qy    373 ProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThr 392
Db   1108 GTGGTCATTCCTAGTGATTATTTTCAGATAGTG---GGAGGAAGCAACTTACGGATACTT 1164

Qy    393 GlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThr 412
Db   1165 GGGGTGGTGAAGTCAGATGAAGGCTTTTATCAATGTGTGGCTGAAAATGAGGCTGGAAAT 1224

Qy    413 AlaCysAlaAlaAlaProLeu---------------AlaValValValArgGluGlyLeu 427
Db   1225 GCCCAGACCAGTGCACAGCTCATTGTCCCTAAGCCTGCAATCCCAAGCTCCAGTGTCCTC 1284

Qy    428 ProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAla 447
Db   1285 CCTTCGGCTCCCAGAGATGTGGTCCCTGTCTTGGTTTCCAGCCGATTTGTCCGTCTCAGC 1344

Qy    448 TrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLys 467
Db   1345 TGGCGCCCACCTGCAGAAGCGAAAGGGAACATTCAAACTTTCACGGTCTTTTTC------ 1398

Qy    468 AlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsn---------AspThr 484
Db   1399 TCCAGAGAAGGTGACAACAGGGAACGA------GCATTGAATACAACACAGCCTGGGTCC 1452

Qy    485 ThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAla 504
Db   1453 CTTCAGCTCACTGTGGGAAACCTGAAGCCAGAAGCCATGTACACCTTTCGAGTTGTGGCT 1512

Qy    505 TyrSerGlnLeuGlyAlaSerArgThrSerSerPro------------------------ 516
Db   1513 TACAATGAATGGGGACCGGGAGAGAGTTCTCAACCCATCAAGGTGGCCACACAGCCTGAG 1572

Qy    516 ------------------------------------------------------------ 516
Db   1573 TTGCAAGTTCCAGGGCCAGTAGAAAACCTGCAAGCTGTATCTACCTCACCTACCTCAATT 1632

Qy    516 ------------------------------------------------------------ 516
Db   1633 CTTATTACCTGGGAACCCCCTGCCTATGCAAACGGTCCAGTCCAAGGTTACAGATTGTTC 1692

Qy    516 ------------------------------------------------------------ 516
Db   1693 TGCACTGAGGTGTCCACAGGAAAAGAACAGAATATAGAGGTTGATGGACTATCTTATAAA 1752

Qy    516 ------------------------------------------------------------ 516
Db   1753 CTGGAAGGCCTGAAAAAATTCACCGAATATAGTCTTCGATTCTTAGCTTATAATCGCTAT 1812

Qy    517 ------------------------AlaLeuValHisThrLeuAspAspValProSerAla 528
Db   1813 GGTCCGGGCGTCTCTACTGATGATATAACAGTGGTTACACTTTCTGACGTGCCAAGTGCC 1872

Qy    529 AlaProGln---LeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeu 547
Db   1873 CCGCCTCAGAACGTCTCCCTGGAAGTGGTCAATTCAAGAAGTATCAAAGTTAGCTGGCTG 1932

Qy    548 ProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeu 567
Db   1933 CCTCCTCCATCAGGAACACAAAATGGATTTATTACCGGCTATAAAATTCGACACAGAAAG 1992

Qy    568 GlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeu 587
Db   1993 ACGACCCGCAGGGGTGAGATGGAAACACTGGAGCCAAACAACCTCTGGTACCTA---TTC 2049

Qy    588 AsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGly 607
```

```
Db  2050 ACAGGACTGGAGAAAGGAAGTCAGTACAGTTTCCAGGTGTCAGCCATGACAGTCAATGGT 2109

Qy   608 TyrGlyValProSerGlnTrpMetGlnHisArgThrPro---GlyValHisAsnGlnSer 626

Db  2110 ACTGGACCACCTTCCAACTGGTATACTGCAGAGACTCCAGAGAATGATCTAGATGAATCT 2169

Qy   627 HisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuValVal 646

Db  2170 CAAGTTCCTGATCAACCAAGCTCTCTTCATGTGAGGCCCCAGACTAACTGCATCATCATG 2229

Qy   647 SerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyrTrpGly 665

Db  2230 AGTTGGACTCCTCCCTTGAACCCAAACATCGTGGTGCGAGGTTATATTATCGGTTATGGC 2289

Qy   666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685

Db  2290 GTTGGGAGCCCTTACGCTGAGACA---------------------------------- 2313

Qy   686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705

Db  2314 ------------------GTGCGTGTGGACAGCAAGCAGCGATATTATTCCATTGAGAGG 2355

Qy   706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725

Db  2356 TTAGAGTCAAGTTCCCATTATGTAATCTCCCTAAAAGCTTTTAACAATGCCGGAGAAGGA 2415

Qy   726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745

Db  2416 GTTCCTCTTTATGAAAGTGCCACCACCAGGTCTATAACCGATCCCACTGACCCAGTTGAT 2475

Qy   746 ---------------------------------------------ArgGlyProProLeu 750

Db  2476 TATTATCCTTTGCTTGATGATTTCCCCACCTCGGTCCCAGATCTCTCCACCCCCATGCTC 2535

Qy   751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrp--- 769

Db  2536 CCACCAGTAGGTGTACAGGCTGTGGCTCTTACCCATGATGCTGTGAGGGTCAGCTGGGCA 2595

Qy   770 ---------------LysLysProAspPheThrThrValLysIleValAsnTyrThrVal 784

Db  2596 GACAACTCTGTCCCTAAGAACCAAAAGACGTCTGAGGTGCGACTT------TACACCGTC 2649

Qy   785 ArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr--------------- 799

Db  2650 CGG---------TGGAGAACCAGCTTTTCTGCAAGTGCAAAATACAAGTCAGAAGACACA 2700

Qy   800 ------------TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPhe 815

Db  2701 ACATCTCTAAGTTACACAGCAACA--------------------GGCCTCAAACCAAAC 2739

Qy   816 ThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySer 835

Db  2740 ACAATGTATGAATTCTCGGTCATGGTAACAAAAAACAGAAGGTCCAGTACTTGGAGCATG 2799

Qy   836 ValValGluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeu 855

Db  2800 ACTGCACATGCCACCACGTATGAAGCAGCCCCCACCTCTGCTCCCAAGGACTTTACAGTC 2859

Qy   856 SerProLeuThr---------------ProSerThrValArgLeuHisTrpCysProProThr 871

Db  2860 ------ATTACTAGGGAAGGGAAGCCTCGTGCCGTCATTGTGAGTTGGCAGCCTCCCTTG 2913

Qy   872 GluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnPro 891

Db  2914 GAAGCCAATGGGAAAATTACTGCTTACATCTTATTTTATACCTTGGACAAGAACATCCCA 2973

Qy   892 GluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGly 911

Db  2974 ATTGATGACTGGATTATGGAAACAATCAGTGGTGATAGGCTTACTCATCAAATCATGGAT 3033

Qy   912 LeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGly 931

Db  3034 CTCAACCTTGATACTATGTATTACTTTCGAATTCAAGCACGAAATTCAAAAGGAGTGGGG 3093

Qy   932 ProPheSer------------------------------------------------- 934

Db  3094 CCACTCTCTGATCCCATCCTCTTCAGGACTCTGAAAGTGGAACACCCTGACAAAATGGCT 3153

Qy   935 ------------------------------------------------ArgLeuGlnAsp 938

Db  3154 AATGACCAAGGTCGTCATGGAGATGGAGGTTATTGGCCAGTTGATACTAATTTGATTGAT 3213

Qy   939 ValIleThrLeuGlnGlu------------------------------------------Thr 945

Db  3214 AGAAGCACCCTAAATGAGCCGCCAATTGGACAAATGCACCCCCCGCATGGCAGTGTCACT 3273

Qy   946 PheSerAspSerLeuAspValHisAlaValThrGlyIleIleValGlyValCysLeuGly 965

Db  3274 CCTCAGAAGAACAGCAACCTGCTTGTGATCATTGTGGTCACCGTTGGTGTCATCACAGTG 3333

Qy   966 LeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla 985

Db  3334 CTGGTAGTGGTCATCGTGGCTGTGATTTGCACCCGACGCTCTTCAGCCCAGCAGAGAAAG 3393

Qy   986 LeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArg 1005

Db  3394 AAACGGGCCACCCACAGTGCTGGCAAAAGGAAGGGC------------AGCCAGAAGGAC 3441

Qy  1006 LeuGlyProProSerValProAlaAlaHis---GluLeuGluSerLeuValHisProArg 1024

Db  3442 CTCCGACCCCCTGATCTTTGGATCCATCATGAAGAAATGGAGATGAAAAATATTGAAAAG 3501

Qy  1025 ProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSerLeu 1044

Db  3502 CCATCTGGCACTGACCCTGCAGGAAGGGAC---------------------------- 3531

Qy  1045 MetGlyGlySerValSerAspCysArg---------------GlyHisSerLys-ArgLy 1059

Db  3532 ------TCTCCCATCCAAAGTTGCCAAGACCTCACACCAGTCAGCCACAGCCAGTCAGAA 3585

Qy  1059 sIleSerTrpAlaGln------------------------------------------Al 1065

Db  3586 ACCCAACTGGGAAGCAAAAGCACCTCTCATTCAGGTCAAGACACTGAGGAAGCAGGGAGC 3645

Qy  1065 aGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGln----------- 1081

Db  3646 TCTATGTCCACTCTGGAGAGGTCGCTGGCTGCACGCCGAGCC-CCCCGGGCCAAGCTCAT 3704

Qy  1082 -------------GlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAl 1097

Db  3705 GATTCCCATGGATGCCCAGTCCAACAATCCTGCTGTCGTGAGCGCCATCCCGGTGCCAAC 3764

Qy  1097 aGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGl 1117

Db  3765 G--------------------CTAGAAAGTGCCCAGTACCCAGGAATC----------- 3792

Qy  1117 yArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSe 1137

Db  3793 -------CTCCCGTCTCCCACCTGTGGA------------------------------ 3813

Qy  1137 rAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGl 1157

Db  3814 --------------------------------------------------TATCCCCACCC 3824

Qy  1157 uProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAl 1177

Db  3825 GCAGTTCACTCTCCGGCCTGTGCCATTCCCAACACTCTCAGTGGACCGAGGTTTCGGAGC 3884

Qy  1177 aAspTrpLeuGlyArgGluLeuGlyGlyCysGln---ProThrThrSerGlyProGluAr 1196

Db  3885 A---------GGAAGAAGTCAGTCAGTGAGTGAAGGACCAACTACCCAACAACCACCTAT 3935

Qy  1196 gLeu---ThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnPr 1215

Db  3936 GCTGCCCCCATCTCAGCCTGAGCATTCTAGCAGCGAGGAGGCACCAAGCAGAACCATCCC 3995
```

```
Qy       1215 oSerThrAlaIleGluGluAla------------------------------------- 1222
Db       3996 CACAGCTTGTGTTCGACCAACTCACCCACTCCGCAGCTTTGCTAATCCTTTGCTACCTCC 4055

Qy       1223 -ProGlyLysSerCysGlnProLysAlaLeuCysProLeuThr---------------Va 1237
Db       4056 ACCAATGAGTGCAATAGAACCGAAAGTC------CCTTACACACCACTTTTGTCTCAGCC 4109

Qy       1237 lSerProSerLeuProArgAlaProValSerSerAlaGlnVal 1251
Db       4110 AGGGCCCACTCTTCCTAAGACCCATGTGAAAACAGCCTCCCTT 4152
```

RESULT 13
PCT-US94-05277-1

```
; Sequence 1, Application PC/TUS9405277
;  GENERAL INFORMATION:
;    APPLICANT:  Bruskin, Arthur
;    APPLICANT:  Jarosz, David E.
;    APPLICANT:  Johnson, Karen
;    APPLICANT:  Kinzler, Kenneth W.
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Zabrecky, James R.
;    TITLE OF INVENTION:  Antibodies Specific for DCC Gene Product
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05277
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.42709
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202.508.9100
;      TELEFAX:  202.508.9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4608 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  18q21
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..4342
PCT-US94-05277-1
```

Alignment Scores:

Pred. No.:	8.57e-53	Length:	4608
Score:	943.00	Matches:	357
Percent Similarity:	36.29%	Conservative:	171
Best Local Similarity:	24.54%	Mismatches:	547
Query Match:	14.28%	Indels:	381
DB:	5	Gaps:	47

US-09-754-997A-2 (1-1252) x PCT-US94-05277-1 (1-4608)

```
Qy         51 ValValLeuAspCysThrLeuGlyAlaThrAlaAlaGlyProProThrArgValThrTrp 70
Db        169 GTCCTCCTCGACTGCTCC---GCGGAGTCCGACCGAGGAGTTCCAGTG---ATCAAGTGG 222

Qy         71 SerLysAspGlyAspThrValLeuGluHisGluAsnLeuHisLeu--------------- 85
Db        223 AAGAAAGATGGC-----------------------ATTCATCTGGCCTTGGGAATGGAT 258

Qy         86 ---------------LeuProAsnGlySerLeuTrpLeuSerSerProLeuGluGlnGlu 100
Db        259 GAAAGGAAGCAGCAACTTTCAAATGGGTCTCTGCTGATACAAAACATACTTCATTCC--- 315

Qy        101 AspSerAspAspGluGluAlaLeuArgIleTrpLysValThrGluGlySerTyrSerCys 120
Db        316 ------------------------AGACACCACAAGCCAGATGAGGGACTTTACCAATGT 351

Qy        121 LeuAlaHis---SerProLeuGlyValValAlaSerGlnValAlaValValLysLeuAla 139
Db        352 GAGGCATCTTTAGGAGATTCTGGCTCAATTATTAGTCGGACAGCAAAAGTTGCAGTAGCA 411

Qy        140 ThrLeuGluAspPheSerLeuHisProGluSerGlnIleValGluGluAsnGlyThrAla 159
Db        412 GGACCACTGAGGTTCCTTTCACAGACAGAATCTGTCACAGCCTTCATGGGAGACACAGTG 471

Qy        160 ArgPheGluCysHisThrLysGlyLeuProAlaProIleIleThrTrpGluLysAspGln 179
Db        472 CTACTCAAGTGTGAAGTCATTGGGGAGCCCATGCCAACAATCCACTGGCAGAAGAACCAA 531

Qy        180 ---------ValThrValProGluGluProArgLeuIleThrLeuProLysTrpLeuLeu 196
Db        532 CAAGACCTGACTCCAATCCCAGGTGACTCCCGAGTGGTGGTCTTGCCCTCTGGAGCATTG 591

Qy        197 GlnIleLeuAspValGlnAspSerAspAlaGlySerTyrArgCysValAlaThrAsnSer 216
Db        592 CAGATCAGCCGACTCCAACCGGGGGACATTGGAATTTACCGATGCTCAGCTCGAAATCCA 651

Qy        217 AlaArgGlnArgPheSerGlnGluAlaSerLeuThrValAlaLeuArgGlySerLeuGlu 236
Db        652 GCCAGCTCAAGAACAGGAAATGAAGCAGAAGTCAGAATTTTATCAGATCCAGGACTGCAT 711

Qy        237 AlaThrArgGlyGlnAspValValIleValAlaAlaProGluAsnThrThrValValSer 256
Db        712 ------------AGACAGCTGTATTTTCTGCAAAGACCATCCAATGTAGTAGCCATTGAA 759

Qy        257 GlyGlnAsnValValMetGluCysValAlaSerAlaAspProThrProPheValSerTrp 276
Db        760 GGAAAAGATGCTGTCCTGGAATGTTGTGTTTCTGGCTATCCTCCACCAAGTTTTACCTGG 819

Qy        277 ValArgGlnAspGlyLysProIle------------SerThrAspValIleValLeuGly 292
Db        820 TTACGA------GGCGAGGAAGTCATCCAACTCAGGTCTAAAAAGTATTCTTTATTGGGT 873

Qy        293 ArgThrAsnLeuLeuIleAlaSerAlaGlnProArgHisSerGlyValTyrValCysArg 312
Db        874 GGAAGCAACTTGCTTATCTCCAATGTGACAGATGATGACAGTGGAATGTATACCTGTGTT 933

Qy        313 AlaAsnLysProLeuThrArgAspPheAlaThrAlaAlaGluLeuArgValLeuAla 332
Db        934 GTCACA------TATAAAAATGAGAATATTAGTGCCTCTGCAGAGCTCACAGTCTTGGTT 987

Qy        333 AlaProAlaIleSerGlnAlaProGluAlaLeuSerArgThrArgAlaSerThrAlaArg 352
Db        988 CCGCCATGGTTTTTAAATCATCCTTCCAACCTGTATGCCTATGAAAGCATGGATATTGAG 1047

Qy        353 PheValCysArgAlaSerGlyGluProArgProAlaLeuHisTrpLeuHisAspGlyIle 372
Db       1048 TTTGAATGTACAGTCTCTGGAAAGCCTGTGCCCACTGTGAATTGGATGAAGAATGGAGAT 1107

Qy        373 ProLeuArgProAsnGlyArgValLysValGlnGlyGlyGlyGlySerLeuValIleThr 392
```

```
Db   1108 GTGGTCATTCCTAGTGATTATTTTCAGATAGTG---GGAGGAAGCAACTTACGGATACTT 1164
Qy    393 GlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCysValAlaGluAsnSerAlaGlyThr 412
Db   1165 GGGGTGGTGAAGTCAGATGAAGGCTTTTATCAATGTGTGGCTGAAAATGAGGCTGGAAAT 1224
Qy    413 AlaCysAlaAlaAlaProLeu---------------AlaValValValArgGluGlyLeu 427
Db   1225 GCCCAGACCAGTGCACAGCTCATTGTCCCTAAGCCTGCAATCCCAAGCTCCAGTGTCCTC 1284
Qy    428 ProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerValLeuValAla 447
Db   1285 CCTTCGGCTCCCAGAGATGTGGTCCCTGTCTTGGTTTCCAGCCGATTTGTCCGTCTCAGC 1344
Qy    448 TrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHisTyrGlnLys 467
Db   1345 TGGCGCCCACCTGCAGAAGCGAAAGGGAACATTCAAACTTTCACGGTCTTTTTC------ 1398
Qy    468 AlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsn---------AspThr 484
Db   1399 TCCAGAGAAGGTGACAACAGGGAACGA------GCATTGAATACAACACAGCCTGGGTCC 1452
Qy    485 ThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAla 504
Db   1453 CTTCAGCTCACTGTGGGAAACCTGAAGCCAGAAGCCATGTACACCTTTCGAGTTGTGGCT 1512
Qy    505 TyrSerGlnLeuGlyAlaSerArgThrSerSerPro------------------------ 516
Db   1513 TACAATGAATGGGGACCGGGAGAGAGTTCTCAACCCATCAAGGTGGCCACACAGCCTGAG 1572
Qy    516 ------------------------------------------------------------ 516
Db   1573 TTGCAAGTTCCAGGGCCAGTAGAAAACCTGCAAGCTGTATCTACCTCACCTACCTCAATT 1632
Qy    516 ------------------------------------------------------------ 516
Db   1633 CTTATTACCTGGGAACCCCCTGCCTATGCAAACGGTCCAGTCCAAGGTTACAGATTGTTC 1692
Qy    516 ------------------------------------------------------------ 516
Db   1693 TGCACTGAGGTGTCCACAGGAAAAGAACAGAATATAGAGGTTGATGGACTATCTTATAAA 1752
Qy    516 ------------------------------------------------------------ 516
Db   1753 CTGGAAGGCCTGAAAAAATTCACCGAATATAGTCTTCGATTCTTAGCTTATAATCGCTAT 1812
Qy    517 ------------------------AlaLeuValHisThrLeuAspAspValProSerAla 528
Db   1813 GGTCCGGGCGTCTCTACTGATGATATAACAGTGGTTACACTTTCTGACGTGCCAAGTGCC 1872
Qy    529 AlaProGln---LeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeu 547
Db   1873 CCGCCTCAGAACGTCTCCCTGGAAGTGGTCAATTCAAGAAGTATCAAAGTTAGCTGGCTG 1932
Qy    548 ProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeu 567
Db   1933 CCTCCTCCATCAGGAACACAAAATGGATTTATTACCGGCTATAAAATTCGACACAGAAAG 1992
Qy    568 GlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeu 587
Db   1993 ACGACCCGCAGGGGTGAGATGGAAACACTGGAGCCAAACAACCTCTGGTACCTA---TTC 2049
Qy    588 AsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGly 607
Db   2050 ACAGGACTGGAGAAAGGAAGTCAGTACAGTTTCCAGGTGTCAGCCATGACAGTCAATGGT 2109
Qy    608 TyrGlyValProSerGlnTrpMetGlnHisArgThrPro---GlyValHisAsnGlnSer 626
Db   2110 ACTGGACCACCTTCCAACTGGTATACTGCAGAGACTCCAGAGAATGATCTAGATGAATCT 2169
Qy    627 HisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeuValVal 646
```

```
Db   2170 CAAGTTCCTGATCAACCAAGCTCTCTTCATGTGAGGCCCCAGACTAACTGCATCATCATG 2229
Qy    647 SerTrpGlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyrTrpGly 665
Db   2230 AGTTGGACTCCTCCCTTGAACCCAAACATCGTGGTGCGAGGTTATATTATCGGTTATGGC 2289
Qy    666 GluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGln 685
Db   2290 GTTGGGAGCCCTTACGCTGAGACA------------------------------------ 2313
Qy    686 AlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGln 705
Db   2314 ------------------GTGCGTGTGGACAGCAAGCAGCGATATTATTCCATTGAGAGG 2355
Qy    706 LeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGly 725
Db   2356 TTAGAGTCAAGTTCCCATTATGTAATCTCCCTAAAAGCTTTTAACAATGCCGGAGAAGGA 2415
Qy    726 TyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGln 745
Db   2416 GTTCCTCTTTATGAAAGTGCCACCACCAGGTCTATAACCGATCCCACTGACCCAGTTGAT 2475
Qy    746 ---------------------------------------------ArgGlyProProLeu 750
Db   2476 TATTATCCTTTGCTTGATGATTTCCCCACCTCGGTCCCAGATCTCTCCACCCCCATGCTC 2535
Qy    751 ProProAlaHisValHisAlaGluSerAsnSerSerThrSerIleTrpLeuArgTrp--- 769
Db   2536 CCACCAGTAGGTGTACAGGCTGTGGCTCTTACCCATGATGCTGTGAGGGTCAGCTGGGCA 2595
Qy    770 ---------------LysLysProAspPheThrThrValLysIleValAsnTyrThrVal 784
Db   2596 GACAACTCTGTCCCTAAGAACCAAAAGACGTCTGAGGTGCGACTT------TACACCGTC 2649
Qy    785 ArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyr--------------- 799
Db   2650 CGG---------TGGAGAACCAGCTTTTCTGCAAGTGCAAAATACAAGTCAGAAGACACA 2700
Qy    800 ------------TyrThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPhe 815
Db   2701 ACATCTCTAAGTTACACAGCAACA---------------------GGCCTCAAACCAAAC 2739
Qy    816 ThrLysTyrGluPheAlaValGlnSerHisGlyValAspMetAspGlyProPheGlySer 835
Db   2740 ACAATGTATGAATTCTCGGTCATGGTAACAAAAAACAGAAGGTCCAGTACTTGGAGCATG 2799
Qy    836 ValValGluArgSerThrLeuProAspArgProSerThrProProSerAspLeuArgLeu 855
Db   2800 ACTGCACATGCCACCACGTATGAAGCAGCCCCCACCTCTGCTCCCAAGGACTTTACAGTC 2859
Qy    856 SerProLeuThr------------ProSerThrValArgLeuHisTrpCysProProThr 871
Db   2860 ------ATTACTAGGGAAGGGAAGCCTCGTGCCGTCATTGTGAGTTGGCAGCCTCCCTTG 2913
Qy    872 GluProAsnGlyGluIleValGluTyrLeuIleLeuTyrSerAsnAsnHisThrGlnPro 891
Db   2914 GAAGCCAATGGGAAAATTACTGCTTACATCTTATTTTATACCTTGGACAAGAACATCCCA 2973
Qy    892 GluHisGlnTrpThrLeuLeuThrThrGluGlyAsnIlePheSerAlaGluValHisGly 911
Db   2974 ATTGATGACTGGATTATGGAAACAATCAGTGGTGATAGGCTTACTCATCAAATCATGGAT 3033
Qy    912 LeuGluSerAspThrArgTyrPhePheLysMetGlyAlaArgThrGluValGlyProGly 931
Db   3034 CTCAACCTTGATACTATGTATTACTTTCGAATTCAAGCACGAAATTCAAAAGGAGTGGGG 3093
Qy    932 ProPheSer--------------------------------------------------- 934
Db   3094 CCACTCTCTGATCCCATCCTCTTCAGGACTCTGAAAGTGGAACACCCTGACAAAATGGCT 3153
Qy    935 ---------------------------------------------ArgLeuGlnAsp 938
Db   3154 AATGACCAAGGTCGTCATGGAGATGGAGGTTATTGGCCAGTTGATACTAATTTGATTGAT 3213
```

```
Qy        939 ValIleThrLeuGlnGlu---------------------------------------Thr 945
                  |||||  |||                                         |||
Db       3214 AGAAGCACCCTAAATGAGCCGCCAATTGGACAAATGCACCCCCCGCATGGCAGTGTCACT 3273

Qy        946 PheSerAspSerLeuAspValHisAlaValThrGlyIleIleValGlyValCysLeuGly 965
                 :::   :::::           :::  ||||||||
Db       3274 CCTCAGAAGAACAGCAACCTGCTTGTGATCATTGTGGTCACCGTTGGTGTCATCACAGTG 3333

Qy        966 LeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArgGluAla 985
              |||:::  ::::::    :::       |||  |||:::   :::
Db       3334 CTGGTAGTGGTCATCGTGGCTGTGATTTGCACCCGACGCTCTTCAGCCCAGCAGAGAAAG 3393

Qy        986 LeuProGlyLeuSerSerSerGlyThrProGlyAsnProAlaLeuTyrThrArgAlaArg 1005
                          |||:::|||                         ::::::
Db       3394 AAACGGGCCACCCACAGTGCTGGCAAAAGGAAGGGC------------AGCCAGAAGGAC 3441

Qy       1006 LeuGlyProProSerValProAlaAlaHis---GluLeuGluSerLeuValHisProArg 1024
              |||   ||||||  :::            |||   |||:::|||              :::
Db       3442 CTCCGACCCCCTGATCTTTGGATCCATCATGAAGAAATGGAGATGAAAAAATATTGAAAAG 3501

Qy       1025 ProGlnAspTrpSerProProProSerAspValGluAspLysAlaGluValHisSerLeu 1044
              |||         |||      |||
Db       3502 CCATCTGGCACTGACCCTGCAGGAAGGGAC------------------------------ 3531

Qy       1045 MetGlyGlySerValSerAspCysArg---------------GlyHisSerLys-ArgLy 1059
                          :::   |||:::                |||||||:::  :::||
Db       3532 ------TCTCCCATCCAAAGTTGCCAAGACCTCACACCAGTCAGCCACAGCCAGTCAGAA 3585

Qy       1059 sIleSerTrpAlaGln----------------------------------------Al 1065
              |   :::|||                                               ||
Db       3586 ACCCAACTGGGAAGCAAAAGCACCTCTCATTCAGGTCAAGACACTGAGGAAGCAGGGAGC 3645

Qy       1065 aGlyGlyProAsnTrpAlaGlySerTrpAlaGlyCysGluLeuProGln----------- 1081
              |     |||   |||   |||   |||         |||   |||:::
Db       3646 TCTATGTCCACTCTGGAGAGGTCGCTGGCTGCACGCCGAGCC-CCCCGGGCCAAGCTCAT 3704

Qy       1082 -------------GlySerGlyProArgProAlaLeuThrArgAlaLeuLeuProProAl 1097
                                      ||||||:::      |||:::        |||
Db       3705 GATTCCCATGGATGCCCAGTCCAACAATCCTGCTGTCGTGAGCGCCATCCCGGTGCCAAC 3764

Qy       1097 aGlyThrGlyGlnThrLeuLeuLeuGlnAlaLeuValTyrAspGlyIleLysSerAsnGl 1117
                                 |||:::::    |||  ||||||
Db       3765 G-------------------CTAGAAAGTGCCCAGTACCCAGGAATC----------- 3792

Qy       1117 yArgLysLysProSerProAlaCysArgAsnGlnValGluAlaGluValIleValHisSe 1137
                      |||||||||  |||
Db       3793 -------CTCCCGTCTCCCACCTGTGGA------------------------------ 3813

Qy       1137 rAspPheGlyAlaSerLysGlyCysProAspLeuHisLeuGlnAspLeuGluProGluGl 1157
                                                                  |||
Db       3814 ---------------------------------------------------TATCCCCACCC 3824

Qy       1157 uProLeuThrAlaGluThrLeuProSerThrSerGlyAlaValAspLeuSerGlnGlyAl 1177
                    |||          :::|||      :::   :::|||||         |||||
Db       3825 GCAGTTCACTCTCCGGCCTGTGCCATTCCCAACACTCTCAGTGGACCGAGGTTTCGGAGC 3884

Qy       1177 aAspTrpLeuGlyArgGluLeuGlyGlyCysGln---ProThrThrSerGlyProGluAr 1196
              |         ||||||             :::   |||||||||      |||
Db       3885 A---------GGAAGAAGTCAGTCAGTGAGTGAAGGACCAACTACCCAACAACCACCTAT 3935

Qy       1196 gLeu---ThrCysLeuProGluAlaAlaSerAlaSerCysSerCysSerAspLeuGlnPr 1215
               |||         ||||||   :::|||:::       :::   |||          ||
Db       3936 GCTGCCCCCATCTCAGCCTGAGCATTCTAGCAGCGAGGAGGCACCAAGCAGAACCATCCC 3995

Qy       1215 oSerThrAlaIleGluGluAla-------------------------------------- 1222
              |:::      :::
Db       3996 CACAGCTTGTGTTCGACCAACTCACCCACTCCGCAGCTTTGCTAATCCTTTGCTACCTCC 4055

Qy       1223 -ProGlyLysSerCysGlnProLysAlaLeuCysProLeuThr---------------Va 1237
               |||      :::   :::||||||         |||   |||
Db       4056 ACCAATGAGTGCAATAGAACCGAAAGTC------CCTTACACACCACTTTTGTCTCAGCC 4109

Qy       1237 lSerProSerLeuProArgAlaProValSerSerAlaGlnVal 1251
               |||:::||||||:::      |||   :::|||   :::
Db       4110 AGGGCCCACTCTTCCTAAGACCCATGTGAAAACAGCCTCCCTT 4152
```

RESULT 14

```
US-09-877-730-23
; Sequence 23, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Abuin, Alejandro
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
;  FILE REFERENCE: LEX-0189-USA
;  CURRENT APPLICATION NUMBER: US/09/877,730
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/210,607
;  PRIOR FILING DATE: 2000-06-09
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 1875
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-877-730-23

Alignment Scores:
Pred. No.:               4.35e-51      Length:          1875
Score:                   909.00        Matches:         204
Percent Similarity:      52.83%        Conservative:    95
Best Local Similarity:   36.04%        Mismatches:      205
Query Match:             13.76%        Indels:          62
DB:                      4             Gaps:            9

US-09-754-997A-2 (1-1252) x US-09-877-730-23 (1-1875)

Qy        470 GlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThrThrGluLeuGlnVal 489
              |||::::::|||   |||||||||         :::   |||||||||||       :::
Db         46 GGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACAACTCATTATATTATT 105

Qy        490 ArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAlaTyrSerGlnLeuGly 509
              ||||||||||||   :::::::|||   ||||||:::|||||||||        :::|||
Db        106 GATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCATATATGCCAATGGGA 165

Qy        510 AlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAspValProSerAlaAla 529
              ||||||:::   |||         :::||||||:::|||||||||
Db        166 GCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGATGTTCCCCTGAGACCT 225

Qy        530 ProGlnLeuThrLeuSerSerProAsnProSerAspIleArgValAlaTrpLeuProLeu 549
              |||:::::::::|||:::|||   :::|||:::||||||   ::::::|||||||||:::
Db        226 CCTGAAATTAGTTTGACAAGTCGAAGTCCCACTGATATTCTCATCTCCTGGCTGCCAATC 285

Qy        550 ProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGluTyrGlyLeuGlyLys 569
              |||:::            |||||||||:::   |||::::::   :::   |||
Db        286 CCAGCCAAATATCGGCGGGGCCAAGTGGTGCTGTATCGCTTGTCTTTCCGCCTAAGTACT 345

Qy        570 GluAspGlnValPheSerThrGluValProGlyAsnGluThrGlnLeuThrLeuAsnSer 589
              |||:::   :::            |||:::|||||         :::      |||
Db        346 GAGAATTCAATCCAAGTTCTGGAGCTCCCGGGGACCACGCATGAGTACCTTTTGGAAGGC 405

Qy        590 LeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThrGlyAlaGlyTyrGly 609
              |||:::|||:::      ||||||   |||||||||::::|||      |||   |||
Db        406 CTGAAACCTGACAGTGTCTACCTGGTTCGGATTACTGCTGCCACCAGAGTGGGGCTGGGA 465

Qy        610 ValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsnGlnSerHisValPro 629
                    |||   |||      |||||||||||                        |||
Db        466 GAGTCATCAGTATGGACTTCACATAGGACGCCCAAAGCT---ACAAGCGTGAAAGCCCCT 522

Qy        630 PheAlaProAlaGluLeuLysValArg---AlaLysMetGluSerLeuValValSerTrp 648
```

```
                   :::|||   ||||||  :::                 ::::::  |||   |||
Db        523 AAGTCTCCA---GAGTTGCATTTGGAGCCTCTGAACTGTACCACCATTTCTGTGAGGTGG 579

Qy        649 GlnProProProHisProThr---GlnIleSerGlyTyrLysLeuTyrTrpGlyGluVal 667
              |||            |||     |||   |||||||||||||||:::  |||
Db        580 CAGCAAGATGTAGAGGACACAGCTGCTATTCAGGGCTACAAGCTGTACTACAAGGAAGAA 639

Qy        668 GlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAspGlnAlaTrp 687
               |||   :::|||
Db        640 GGGCAGCAGGAGAAT--------------------------------------------- 654

Qy        688 AspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThrGlnLeuVal 707
                    ||||||:::   |||      |||         |||  |||:::   |||
Db        655 ------GGGCCCATTTTCTTGGATACCAAGGACCTACTCTATACTCTCAGTGGCTTAGAC 708

Qy        708 ProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAspGlyTyrAla 727
              |||   |||   |||   |||:::|||:::|||:::|||       :::|||||||||
Db        709 CCCAGAAGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGATGGCTATCAG 768

Qy        728 AlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIleGlnArgGly 747
              |||               :::           ||||||            |||
Db        769 GCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCTGTTCGTGAT 813

Qy        748 ------------ProProLeuProProAlaHisValHisAlaGluSerAsnSerSerThr 763
                          |||||| |  ||||||   |||::::::|||::::::|||:::|||:::
Db        814 CGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAACACCTCATCT 873

Qy        764 SerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleValAsnTyrThr 783
              ||||||:::|||   |||::::::|||   ||||||       :::|||:::||||||||
Db        874 TCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATTAACTACACC 933

Qy        784 ValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyrThrSerSer 803
              :::|||      |||   ||||||:::|||||||||||||   |||      :::|||
Db        934 ATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTTCAAACATCA 993

Qy        804 GlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPheAlaValGln 823
                   :::|||:::   ||||||:::|||   ||||||||||||||||||||:::
Db        994 GAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTTGCCGTTCGA 1053

Qy        824 SerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSerThrLeuPro 843
                 |||         :::   |||:::      ||||||      |||||||||||
Db       1054 TTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCTACTCTTCCA 1113

Qy        844 AspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrProSerThrVal 863
              :::   |||:::   ||||||      ::::::::::::   :::               |||
Db       1114 GAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGATGACACTGCC 1173

Qy        864 ArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyrLeuIleLeu 883
                 :::   |||   ||||||      |||         :::         |||   ||||||
Db       1174 CTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTATACTATCTTA 1233

Qy        884 TyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThrGluGlyAsn 903
              |||::::::             :::|||   :::|||      ||||||
Db       1234 TATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGTGAAGGGGCA 1293

Qy        904 IlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePheLysMetGly 923
              |||      |||   :::      |||   :::         |||      |||:::
Db       1294 ATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCT 1353

Qy        924 AlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIleThrLeuGln 943
              |||      |||||||||  |||||||||||||          ::::::      |||
Db       1354 GCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCAGTACTTCCA 1413

Qy        944 GluThrPheSerAspSer------------------------------------------ 949
              :::         |||:::|||
Db       1414 AAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTT 1473

Qy        950 ------------------LeuAspValHisAlaValThrGlyIleIleValGlyValCys 963
                                |||||      ::::::|||||||||   |||||||||
Db       1*74 TATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTAGGTGTTGGC 1533

Qy        964 LeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSerSerHisArg 983
              :::    |||    |||:::|||   |||:::        :::    :::|||      |||

Db       1534 ATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGTAAAGCCAGG 1593

Qy        984 GluAla-----------------------------------LeuProGlyLeuSerSer 991
              ::::::                                   ||||||    |||:::
Db       1594 AAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGTACCAGTGCC 1653

Qy        992 SerGlyThrProGlyAsn 997
              |||         ||||||
Db       1654 TCCTTAGCTAGTGGAAAT 1671
```

RESULT 15
US-09-877-730-25
; Sequence 25, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-25

Alignment Scores:			
Pred. No.:	9.49e-43	Length:	1644
Score:	781.50	Matches:	189
Percent Similarity:	41.11%	Conservative:	86
Best Local Similarity:	28.25%	Mismatches:	169
Query Match:	11.83%	Indels:	225
DB:	4	Gaps:	6

US-09-754-997A-2 (1-1252) x US-09-877-730-25 (1-1644)

```
Qy        365 LeuHisTrpLeuHisAspGlyIleProLeuArgProAsnGlyArgValLysValGlnGly 384
              :::    ||||||    :::|||       :::.        |||||||||:::|||:::
Db          1 ATGTCATGGTTGAAAAATGGAAGGAAGATACATTCGAATGGTAGAATTAAAATGTAC--- 57

Qy        385 GlyGlyGlySerLeuValIleThrGlnIleGlyLeuGlnAspAlaGlyTyrTyrGlnCys 404
                          |||||||||    ||||||         :::|||      |||||||||
Db         58 ---AACAGTAAATTGGTAATTAACCAGATTATTCCTGAAGATGATGCTATTTATCAGTGC 114

Qy        405 ValAlaGluAsnSerAlaGlyThrAlaCysAlaAlaAlaProLeuAlaValValValArg 424
              :::||||||||||||   |||:::         :::   |||   |||   ||||||:::
Db        115 ATGGCTGAGAATAGCCAAGGATCTATTTATCTAGAGCCAGACTGACTGTAGTGATGTCA 174

Qy        425 GluGlyLeuProSerAlaProThrArgValThrAlaThrProLeuSerSerSerSerVal 444
              |||      ||||||||||||      |||   |||        :::|||||||||::::::
Db        175 GAAGACAGACCCAGTGCTCCCTATAATGTACATGCTGAAACCATGTCAAGCTCAGCCATT 234

Qy        445 LeuValAlaTrpGluArgProGluLeuHisSerGluGlnIleIleGlyPheSerLeuHis 464
              |||:::|||||||||||||||       :::|||:::::::::|||   :::|||:::|||
Db        235 CTTTTAGCCTGGGAGAGGCCACTTTATAATTCAGACAAAGTCATTGCCTATTCTGTACAC 294

Qy        465 TyrGlnLysAlaArgGlyValAspAsnValGluTyrGlnPheAlaValAsnAsnAspThr 484
              |||   ||||||   |||:::::::|||   |||||||||         :::   |||||||||
Db        295 TACATGAAAGCAGAAGGTTTAAATAATGAAGAGTATCAAGTAGTCATCGGAAATGACACA 354

Qy        485 ThrGluLeuGlnValArgAspLeuGluProAsnThrAspTyrGluPheTyrValValAla 504
              |||         :::   ||||||||||||     ::::::|||   ||||||:::||||||
Db        355 ACTCATTATATTATTGATGACTTAGAGCCTGCCAGCAATTATACTTTCTACATTGTAGCA 414
```

```
Qy        505 TyrSerGlnLeuGlyAlaSerArgThrSerSerProAlaLeuValHisThrLeuAspAsp 524
              |||      :::|||||||||:::   |||                 :::||||||:::|||
Db        415 TATATGCCAATGGGAGCCAGCCAGATGTCTGACCATGTGACACAGAATACTCTAGAGGAT 474

Qy        525 ValProSerAlaAlaProGlnLeuThrLeuSerSerProAsnProSerAspIleArgVal 544
                    |||
Db        475 GACCCCAGA--------------------------------------------------- 483

Qy        545 AlaTrpLeuProLeuProSerSerLeuSerAsnGlyGlnValLeuLysTyrLysIleGlu 564

Db        483 ------------------------------------------------------------ 483

Qy        565 TyrGlyLeuGlyLysGluAspGlnValPheSerThrGluValProGlyAsnGluThrGln 584

Db        483 ------------------------------------------------------------ 483

Qy        585 LeuThrLeuAsnSerLeuGlnProAsnLysValTyrArgValArgIleSerAlaGlyThr 604

Db        483 ------------------------------------------------------------ 483

Qy        605 GlyAlaGlyTyrGlyValProSerGlnTrpMetGlnHisArgThrProGlyValHisAsn 624

Db        483 ------------------------------------------------------------ 483

Qy        625 GlnSerHisValProPheAlaProAlaGluLeuLysValArgAlaLysMetGluSerLeu 644

Db        483 ------------------------------------------------------------ 483

Qy        645 ValValSerTrpGlnProProProHisProThrGlnIleSerGlyTyrLysLeuTyrTrp 664

Db        483 ------------------------------------------------------------ 483

Qy        665 GlyGluValGlyThrGluGluGluAlaAspGlyAspArgProProGlyGlyArgGlyAsp 684

Db        483 ------------------------------------------------------------ 483

Qy        685 GlnAlaTrpAspValGlyProValArgLeuLysLysLysValLysGlnTyrGluLeuThr 704

Db        483 ------------------------------------------------------------ 483

Qy        705 GlnLeuValProGlyArgProTyrGluValLysLeuValAlaPheAsnLysHisGluAsp 724
                                 |||   |||   |||:::|||:::|||:::|||      :::|||
Db        484 ---------------AGAAAATATCATGTGAGACTCCTGGCTTACAACAACATAGACGAT 528

Qy        725 GlyTyrAlaAlaValTrpLysGlyLysThrGluLysAlaProThrProAspLeuProIle 744
              ||||||   |||                  :::         ||||||
Db        529 GGCTATCAGGCA---------------GATCAGACTGTCAGCACTCCAGGATGCGTGTCT 573

Qy        745 GlnArgGly------------ProProLeuProProAlaHisValHisAlaGluSerAsn 760
                   |||               ||||||   ||||||   |||::::::|||:::::::|||
Db        574 GTTCGTGATCGCATGGTCCCTCCTCCACCACCACCCCACCATCTCTATGCGAAGGCTAAC 633

Qy        761 SerSerThrSerIleTrpLeuArgTrpLysLysProAspPheThrThrValLysIleVal 780
              :::|||:::||||||:::|||   |||::::::|||   ||||||      :::|||:::
Db        634 ACCTCATCTTCCATCTTCCTGCACTGGAGGAGGCCTGCATTCACCGCTGCACAAATCATT 693

Qy        781 AsnTyrThrValArgPheGlyProTrpGlyLeuArgAsnAlaSerLeuValThrTyrTyr 800
              ||||||||:::|||      |||   ||||||:::||||||||||||||||   |||
Db        694 AACTACACCATCCGCTGTAATCCTGTTGGCCTGCAGAATGCTTCTTTGGTTCTGTACCTT 753

Qy        801 ThrSerSerGlyGluAspIleLeuIleGlyGlyLeuLysProPheThrLysTyrGluPhe 820
                 :::|||      :::|||:::   ||||||:::|||   |||||||||||||||
Db        754 CAAACATCAGAAACTCACATGTTGGTTCAAGGTCTAGAACCAAACACCAAATACGAATTT 813

Qy        821 AlaValGlnSerHisGlyValAspMetAspGlyProPheGlySerValValGluArgSer 840
              ||||||:::   |||         :::      |||:::      ||||||      |||
Db        814 GCCGTTCGATTACATGTGGATCAGCTTTCCAGTCCTTGGAGCCCTGTAGTCTACCATTCT 873

Qy        841 ThrLeuProAspArgProSerThrProProSerAspLeuArgLeuSerProLeuThrPro 860
              |||||||||:::   |||:::   ||||||   ::::::::::::   :::
Db        874 ACTCTTCCAGAAGCACCAGCAGGCCCACCAGTTGGAGTAAAAGTGACATTAATAGAGGAT 933

Qy        861 SerThrValArgLeuHisTrpCysProProThrGluProAsnGlyGluIleValGluTyr 880
                 |||      :::   |||   ||||||      |||         :::      |||
Db        934 GACACTGCCCTGGTTTCTTGGAAACCCCCTGATGGCCCAGAAACAGTTGTGACCCGCTAT 993

Qy        881 LeuIleLeuTyrSerAsnAsnHisThrGlnProGluHisGlnTrpThrLeuLeuThrThr 900
                 |||||||||::::::                     :::|||      :::|||
Db        994 ACTATCTTATATGCATCTAGGAAGGCCTGGATTGCAGGAGAGTGGCAGGTCTTACACCGT 1053

Qy        901 GluGlyAsnIlePheSerAlaGluValHisGlyLeuGluSerAspThrArgTyrPhePhe 920
              ||||||   |||         |||   :::         |||   :::         |||
Db       1054 GAAGGGGCAATAACCATGGCTTTGCTAGAAAACTTGGTAGCAGGAAATGTGTACATTGTC 1113

Qy        921 LysMetGlyAlaArgThrGluValGlyProGlyProPheSerArgLeuGlnAspValIle 940
              |||:::   |||      |||||||||   ||||||||||||         ::::::
Db       1114 AAGATATCTGCATCCAATGAGGTGGGAGAAGGACCCTTTTCAAATTCTGTGGAGCTGGCA 1173

Qy        941 ThrLeuGlnGluThrPheSerAspSer--------------------------------- 949
                 |||   :::      |||:::|||
Db       1174 GTACTTCCAAAGGAAACCTCTGAATCAAATCAGAGGCCCAAGCGTTTAGATTCTGCTGAT 1233

Qy        950 ---------------------------LeuAspValHisAlaValThrGlyIleIleVal 960
                                         ||||||      ::::::|||||||||   |||
Db       1234 GCCAAAGTTTATTCAGGATATTACCATCTGGACCAAAAATCAATGACTGGCATTGCTGTA 1293

Qy        961 GlyValCysLeuGlyLeuLeuCysLeuLeuAlaCysMetCysAlaGlyLeuArgGlnSer 980
              ||||||   :::   |||   |||:::|||   |||:::         :::   :::|||
Db       1294 GGTGTTGGCATAGCCTTGACCTGCATCCTCATCTGTGTTCTCATCTTGATATACCGAAGT 1353

Qy        981 SerHisArgGluAla------------------------------------LeuProGly 988
                 |||::::::                                       ||||||
Db       1354 AAAGCCAGGAAATCATCTGCTTCCAAGACGGCACAGAATGGAACTCAACAGTTACCTCGT 1413

Qy        989 LeuSerSerSerGlyThrProGlyAsn 997
                 |||:::|||         ||||||
Db       1414 ACCAGTGCCTCCTTAGCTAGTGGAAAT 1440
```

Search completed: October 14, 2003, 09:29:47
Job time : 274 secs

```
OM nucleic - nucleic search, using sw model

Run on:         October 14, 2003, 15:48:26 ; Search time 10461.9 Seconds
                                             (without alignments)
                                             10933.318 Million cell updates/sec

Title:          US-09-754-997A-3
Perfect score:  2796
Sequence:       1 ggggagctgccattgccccca..........cagactccttggatgtgcac 2796

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2796	100.0	2796	6	AX191261	AX191261 Sequence
	2	2796	100.0	3756	6	AX191303	AX191303 Sequence
	3	2796	100.0	6176	6	AX191259	AX191259 Sequence
	4	2796	100.0	6176	10	AF176694	AF176694 Mus muscu
	5	2775.2	99.3	6219	10	AB052620	AB052620 Mus muscu
	6	2762.2	98.8	6222	10	AB052621	AB052621 Mus muscu
	7	2595.2	92.8	6301	10	AK122535	AK122535 Mus muscu
	8	2159.2	77.2	6485	9	AB052622	AB052622 Homo sapi
	9	2090.8	74.8	3741	6	AX418406	AX418406 Sequence
	10	1491.4	53.3	5549	9	AB046848	AB046848 Homo sapi
c	11	415	14.8	116580	10	AC112162	AC112162 Mus Muscu
c	12	415	14.8	158055	2	AC110235	AC110235 Mus muscu
c	13	376.8	13.5	224227	2	AC106509	AC106509 Rattus no
c	14	331.2	11.8	176282	2	AC068507	AC068507 Homo sapi
	15	331.2	11.8	198295	9	AC105129	AC105129 Homo sapi
	16	316.8	11.3	2967	9	BC042054	BC042054 Homo sapi
	17	316.6	11.3	2976	6	AR237558	AR237558 Sequence
	18	316.6	11.3	3210	6	AR237553	AR237553 Sequence
	19	316.6	11.3	3219	6	AR237561	AR237561 Sequence
	20	316.6	11.3	3874	6	AR237568	AR237568 Sequence
	21	316.4	11.3	2700	6	AX497857	AX497857 Sequence
	22	315.4	11.3	3453	6	AR237556	AR237556 Sequence
	23	302.8	10.8	3198	10	BC053057	BC053057 Mus muscu
	24	301.2	10.8	3146	10	AF026465	AF026465 Mus muscu
	25	288	10.3	288	6	AX191277	AX191277 Sequence
	26	287.4	10.3	2729	9	AK095529	AK095529 Homo sapi
c	27	252.6	9.0	115135	10	AC112161	AC112161 Mus Muscu
	28	252	9.0	252	6	AX191281	AX191281 Sequence
	29	249	8.9	249	6	AX191273	AX191273 Sequence
	30	249	8.9	249	6	AX191275	AX191275 Sequence
	31	246	8.8	246	6	AX191279	AX191279 Sequence
	32	243	8.7	243	6	AX191265	AX191265 Sequence
	33	217	7.8	2481	6	AR237560	AR237560 Sequence
	34	217	7.8	2715	6	AR237555	AR237555 Sequence
	35	217	7.8	2724	6	AR237562	AR237562 Sequence
	36	217	7.8	2958	6	AR237557	AR237557 Sequence
c	37	215.8	7.7	147760	9	AC011846	AC011846 Homo sapi
	38	215.8	7.7	176282	2	AC068507	AC068507 Homo sapi
	39	213.8	7.6	2139	6	AR237563	AR237563 Sequence
	40	213.8	7.6	2382	6	AR237566	AR237566 Sequence
	41	195	7.0	195	6	AX191271	AX191271 Sequence
	42	192	6.9	192	6	AX191267	AX191267 Sequence
	43	189	6.8	189	6	AX191269	AX191269 Sequence
c	44	158.8	5.7	256608	2	AC021040	AC021040 Homo sapi
	45	158	5.7	158055	2	AC110235	AC110235 Mus muscu

ALIGNMENTS

```
RESULT 1
AX191261
LOCUS       AX191261                2796 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 3 from Patent WO0149714.
ACCESSION   AX191261
VERSION     AX191261.1  GI:15209512
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 3 12-JUL-2001;
```

```
                     Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2796
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             <1..>2796
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51121.1"
                     /db_xref="GI:15209513"
                     /translation="GELPLPQETTVKLSCDEGPLQVILGPEQAVVLDCTLGATAAGPP
                     TRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRIWKVTEGSYSCLAH
                     SPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGLPAPIITWEKDQVT
                     VPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSARQRFSQEASLTVALRGSLEATR
                     GQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSWVRQDGKPISTDVIVLGRTNLL
                     IASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAPAISQAPEALSRTRASTARFVC
                     RASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGTA
                     CAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQKARGV
                     DNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDVPSAA
                     PQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEDQVFSTEVPGNETQLTL
                     NSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAKMESLV
                     VSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKVKQYEL
                     TQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVHAESNS
                     STSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKPFTKYE
                     FAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEI
                     VEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSR
                     LQDVITLQETFSDSLDVH"
BASE COUNT      591 a    849 c    821 g    535 t
ORIGIN

  Query Match             100.0%;  Score 2796;  DB 6;  Length 2796;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780

Qy        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840

Qy        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900

Qy        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960

Qy        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020

Qy       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Qy       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Qy       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Qy       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Qy       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Qy       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Qy       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Qy       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Qy       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Qy       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Qy       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
```

```
Qy     1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Qy     1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Qy     1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Qy     1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Qy     1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Qy     1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Qy     2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Qy     2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Qy     2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Qy     2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Qy     2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Qy     2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Qy     2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Qy     2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Qy     2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Qy     2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Qy     2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Qy     2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Qy     2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
            ||||||||||||||||||||||||||||||||||||
Db     2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
```

```
RESULT 2
AX191303
LOCUS       AX191303                3756 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 45 from Patent WO0149714.
ACCESSION   AX191303
VERSION     AX191303.1  GI:15209554
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 45 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..3756
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
BASE COUNT      774 a   1167 c   1105 g    710 t
ORIGIN

  Query Match             100.0%;  Score 2796;  DB 6;  Length 3756;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123

Qy       61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183

Qy      121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243

Qy      181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303

Qy      241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363

Qy      301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423

Qy      361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483

Qy      421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543

Qy      481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603

Qy      541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663
```

```
Qy    601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723

Qy    661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783

Qy    721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843

Qy    781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903

Qy    841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963

Qy    901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023

Qy    961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083

Qy   1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143

Qy   1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203

Qy   1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263

Qy   1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323

Qy   1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383

Qy   1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443

Qy   1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503

Qy   1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563

Qy   1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623

Qy   1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683

Qy   1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743

Qy   1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803

Qy   1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863

Qy   1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923

Qy   1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983

Qy   1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043

Qy   1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103

Qy   2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163

Qy   2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223

Qy   2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283

Qy   2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343

Qy   2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403

Qy   2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463

Qy   2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523

Qy   2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583

Qy   2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643

Qy   2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703

Qy   2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763

Qy   2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823

Qy   2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
          ||||||||||||||||||||||||||||||||||||
```

```
Db      2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859


RESULT 3
AX191259
LOCUS       AX191259                6176 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 1 from Patent WO0149714.
ACCESSION   AX191259
VERSION     AX191259.1  GI:15209510
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Salbaum,J.M.
  TITLE     Nope polypeptides, encoding nucleic acids and methods of use
  JOURNAL   Patent: WO 0149714-A 1 12-JUL-2001;
            Neurosciences Research Foundation Inc. (US)
FEATURES             Location/Qualifiers
     source          1..6176
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             1..3759
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51120.1"
                     /db_xref="GI:15209511"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

  Query Match            100.0%;  Score 2796;  DB 6;  Length 6176;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843

Qy        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903

Qy        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963

Qy        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023

Qy        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083

Qy       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143

Qy       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203

Qy       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263

Qy       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323

Qy       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383
```

```
Qy       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443

Qy       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503

Qy       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563

Qy       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623

Qy       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683

Qy       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743

Qy       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803

Qy       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863

Qy       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923

Qy       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983

Qy       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043

Qy       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103

Qy       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163

Qy       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223

Qy       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283

Qy       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343

Qy       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403

Qy       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463

Qy       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523

Qy       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583

Qy       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643

Qy       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703

Qy       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763

Qy       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823

Qy       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
              ||||||||||||||||||||||||||||||||||||
Db       2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859
```

RESULT 4
AF176694
LOCUS AF176694 6176 bp mRNA linear ROD 27-APR-2000
DEFINITION Mus musculus neighbor of Punc e11 protein (Nope) mRNA, complete cds.
ACCESSION AF176694
VERSION AF176694.1 GI:7650185
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 6176)
AUTHORS Salbaum,J.M. and Kappen,C.
TITLE Cloning and expression of nope, a new mouse gene of the immunoglobulin superfamily related to guidance receptors
JOURNAL Genomics 64 (1), 15-23 (2000)
MEDLINE 20175427
PUBMED 10708514
REFERENCE 2 (bases 1 to 6176)
AUTHORS Salbaum,J.M.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1999) The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA
FEATURES Location/Qualifiers
source 1. .6176
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB"
/db_xref="taxon:10090"
/chromosome="9"
/map="band D-E1; 24.59 cR from D9Mit143"
gene 1. .6176
/gene="Nope"
CDS 1. .3759
/gene="Nope"
/note="immunoglobulin superfamily; related to guidance receptors"
/codon_start=1
/product="neighbor of Punc e11 protein"
/protein_id="AAF65930.1"
/db_xref="GI:7650186"
/translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR

```
                    FECHTKGLPAPIITWEKDQVTVPEEPRLITLPKWLLQILDVQDSDAGSYRCVATNSAR
                    QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQNVVMECVASADPTPFVSW
                    VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPLTRDFATAAAELRVLAAP
                    AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                    QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                    PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                    GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                    LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                    QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWGEVGTEEEADGDRPPGGR
                    GDQAWDVGPVRLKKKVKQYELTQLVPGRPYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                    DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                    YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                    PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                    DTRYFFKMGARTEVGPGPFSRLQDVITLQETFSDSLDVHAVTGIIVGVCLGLLCLLAC
                    MCAGLRQSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                    PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                    LTRALLPPAGTGQTLLLQALVYDGIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                    DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                    SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
     misc_feature   1. .23
                    /gene="Nope"
                    /note="unable to sequence region in mRNA"
BASE COUNT     1316 a   1756 c   1634 g   1470 t
ORIGIN

  Query Match             100.0%;  Score 2796;  DB 10;  Length 6176;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2796;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 123

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 183

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 243

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 303

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 363

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 423

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 483

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 543

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 603

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 663

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 723

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 783

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 843

Qy        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 903

Qy        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 963

Qy        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 1023

Qy        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1083

Qy       1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1143

Qy       1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1203

Qy       1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1263

Qy       1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1323

Qy       1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1324 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1383

Qy       1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1443

Qy       1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1444 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1503

Qy       1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1504 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1563

Qy       1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1564 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1623

Qy       1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1624 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1683

Qy       1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1684 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1743

Qy       1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1744 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1803

Qy       1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1804 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1863

Qy       1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1864 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1923

Qy       1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1924 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1983

Qy       1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1984 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2043

Qy       1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2044 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2103

Qy       2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2104 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2163

Qy       2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2164 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2223

Qy       2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2224 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2283

Qy       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2284 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2343

Qy       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2344 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2403

Qy       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2404 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2463

Qy       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2464 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2523

Qy       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2524 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2583

Qy       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2584 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2643

Qy       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2644 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2703

Qy       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2704 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2763

Qy       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2764 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2823

Qy       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
              ||||||||||||||||||||||||||||||||||||
Db       2824 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2859
```

RESULT 5

```
AB052620
LOCUS       AB052620                6219 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36 mRNA, complete cds.
ACCESSION   AB052620
VERSION     AB052620.1  GI:11862938
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6219)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1. .6219
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
                     /clone="DDM36"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)~vector:pDDM36ORF"
     gene            1. .6219
                     /gene="DDM36"
     CDS             50. .3808
                     /gene="DDM36"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36"
                     /protein_id="BAB19278.1"
                     /db_xref="GI:11862939"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHN
                     QSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGR
                     GDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTP
                     DLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVT
                     YYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLS
                     PLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLES
                     DTRYFFKMGARTEVGPGPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLAC
                     MCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSP
                     PPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCP
                     DLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAA
                     SASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1319 a   1772 c   1652 g   1476 t
ORIGIN

  Query Match             99.3%;  Score 2775.2;  DB 10;  Length 6219;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2783;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;
```

```
Qy      1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 172

Qy     61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 232

Qy    121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 292

Qy    181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 352

Qy    241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 412

Qy    301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 472

Qy    361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 532

Qy    421 GAATGCCACACCAAGGGCCTTCCAGCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 GAATGCCACACCAAGGGCCTTCCAGCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 592

Qy    481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
          |||||||||||| |||||||||||||||||||||  |  |||||||||||||||||||||
Db    593 GTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATGTC 652

Qy    541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 712

Qy    601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    713 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 772

Qy    661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    773 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTG 832

Qy    721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 892

Qy    781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    893 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 952

Qy    841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||| |||
Db    953 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTC 1012

Qy    901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1013 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAG 1072

Qy    961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1132

Qy   1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGCGCGTCAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGCGCGTCAAG 1192

Qy   1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   1193 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGCTAC 1252

Qy   1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1312

Qy   1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1372

Qy   1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1373 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1432

Qy   1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1433 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1492

Qy   1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1552

Qy   1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1553 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1612

Qy   1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1613 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1672

Qy   1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1673 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1732

Qy   1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1792

Qy   1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1852

Qy   1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1853 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1912

Qy   1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1913 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1972

Qy   1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1973 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 2032

Qy   1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2033 CTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 2092

Qy   1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2093 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2152

Qy   2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   2153 GAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2212

Qy   2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2213 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2272

Qy   2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2273 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2332

Qy       2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2333 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2392

Qy       2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2393 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2452

Qy       2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2453 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2512

Qy       2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2513 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2572

Qy       2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2573 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2632

Qy       2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2633 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2692

Qy       2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2693 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2752

Qy       2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2753 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2812

Qy       2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2813 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2872

Qy       2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
              |||||| |||||||||||||||||||||||||||||
Db       2873 CTGCAAAAGACATTCTCAGACTCCTTGGATGTGCAC 2908
```

```
RESULT 6
AB052621
LOCUS       AB052621                6222 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Mus musculus DDM36E mRNA, complete cds.
ACCESSION   AB052621
VERSION     AB052621.1  GI:11862940
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6222)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/CBA"
                     /db_xref="taxon:10090"
                     /clone="DDM36E"
                     /cell_line="MMC2"
                     /cell_type="osteoblast"
                     /note="similar to Nope (NM_020043)-vector:pDDM36EORF"
     gene            1..6222
                     /gene="DDM36E"
     CDS             50..3811
                     /gene="DDM36E"
                     /note="similar to Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="DDM36E"
                     /protein_id="BAB19279.1"
                     /db_xref="GI:11862941"
                     /translation="MARADTGRGLLVLTFCLLSARGELPLPQETTVKLSCDEGPLQVI
                     LGPEQAVVLDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDS
                     DDEEALRIWKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTAR
                     FECHTKGLPAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSAR
                     QRFSQEASLTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSW
                     VRQDGKPISTDVIVLGRTNLLIASAQPRHSGVYVCRANKPRTRDFATAAAELRVLAAP
                     AISQAPEALSRTRASTARFVCRASGEPRPALHWLHDGIPLRPNGRVKVQGGGGSLVIT
                     QIGLQDAGYYQCVAENSAGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWER
                     PELHSEQIIGFSLHYQKARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQL
                     GASRTSSPALVHTLDDVPSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYG
                     LGKEEDQVFSTEVPGNETQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVH
                     NQSHVPFAPAELKVRAKMESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGG
                     RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPT
                     PDLPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLV
                     TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                     SPLTPSTVRLHWCPPTEPNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLE
                     SDTRYFFKMGARTEVGPGPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLA
                     CMCAGLRRSSHREALPGLSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWS
                     PPPSDVEDKAEVHSLMGGSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRP
                     ALTRALLPPAGTGQTLLLQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGC
                     PDLHLQDLEPEEPLTAETLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEA
                     ASASCSCSDLQPSTAIEEAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1321 a   1772 c   1653 g   1476 t
ORIGIN

  Query Match            98.8%;  Score 2762.2;  DB 10;  Length 6222;
  Best Local Similarity  99.4%;  Pred. No. 0;
  Matches 2783;  Conservative    0;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 172

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 232

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 292

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 352

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 412

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 472

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 532
```

```
Qy     421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
Db     533 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 592

Qy     481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
Db     593 GTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATGTC 652

Qy     541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
Db     653 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 712

Qy     601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
Db     713 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 772

Qy     661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
Db     773 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTG 832

Qy     721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
Db     833 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 892

Qy     781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
Db     893 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 952

Qy     841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
Db     953 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCGCACGCGTGATTTC 1012

Qy     901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
Db    1013 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCGCCCGAG 1072

Qy     961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
Db    1073 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1132

Qy    1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080
Db    1133 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1192

Qy    1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140
Db    1193 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTTCAGGACGCTGGCTAC 1252

Qy    1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200
Db    1253 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1312

Qy    1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260
Db    1313 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1372

Qy    1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320
Db    1373 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1432

Qy    1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380
Db    1433 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1492

Qy    1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440
Db    1493 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1552

Qy    1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500
Db    1553 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1612

Qy    1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560
Db    1613 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1672

Qy    1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620
Db    1673 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1732

Qy    1621 AAGATAGAGTACGGTTTGGGGAAG---GAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1677
Db    1733 AAGATAGAGTACGGTTTGGGGAAGGAAGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1792

Qy    1678 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1737
Db    1793 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1852

Qy    1738 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1797
Db    1853 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1912

Qy    1798 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1857
Db    1913 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1972

Qy    1858 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1917
Db    1973 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 2032

Qy    1918 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 1977
Db    2033 AAACTCTACTGGAGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 2092

Qy    1978 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2037
Db    2093 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2152

Qy    2038 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2097
Db    2153 TATGAACTGACCCAGTTAGTCCCTGGCAGGCTGTACGAGGTGAAGCTCGTAGCTTTCAAC 2212

Qy    2098 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2157
Db    2213 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2272

Qy    2158 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2217
Db    2273 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2332

Qy    2218 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2277
Db    2333 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2392

Qy    2278 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2337
Db    2393 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2452

Qy    2338 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2397
Db    2453 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2512

Qy    2398 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2457
Db    2513 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2572

Qy    2458 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2517
Db    2573 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2632

Qy    2518 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2577
Db    2633 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2692

Qy    2578 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2637
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2693 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACA 2752

Qy       2638 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2753 GAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTC 2812

Qy       2698 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2813 AAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATT 2872

Qy       2758 ACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796
              |||||||| ||||||||||||||||||||||||||||||
Db       2873 ACTCTGCAAAAGACATTCTCAGACTCCTTGGATGTGCAC 2911
```

RESULT 7
AK122535

```
LOCUS       AK122535                6301 bp    mRNA    linear   ROD 15-MAR-2003
DEFINITION  Mus musculus mRNA for mKIAA1628 protein.
ACCESSION   AK122535
VERSION     AK122535.1  GI:28972801
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
            Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: II. The complete nucleotide sequences of 400 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10, 35-48 (2003)
REFERENCE   2  (bases 1 to 6301)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1..6301
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbg00013"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1..6301
                     /gene="mKIAA1628"
     CDS             <1..3887
                     /gene="mKIAA1628"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=3
                     /evidence=not_experimental
                     /product="mKIAA1628 protein"
                     /protein_id="BAC65817.1"
                     /db_xref="GI:28972802"
                     /translation="EGSQENPSAWHGRGELPLPQETTVKLSCDEGPLQVILGPEQAVV
                     LDCTLGATAAGPPTRVTWSKDGDTVLEHENLHLLPNGSLWLSSPLEQEDSDDEEALRI
                     WKVTEGSYSCLAHSPLGVVASQVAVVKLATLEDFSLHPESQIVEENGTARFECHTKGL
                     PAPIITWEKDQVTVPEESRLITLPNGVLQILDVQDSDAGSYRCVATNSARQRFSQEAS
                     LTVALRGSLEATRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVSWVRQGDQSK
                     KELGIQTSGCDAALQGCSGNLVVGEGVEGTVPASPSSLLLLCHQMESLSPRMSSFWAG
                     PIYSSPARSLGTLESMSAEPTSPARVISPLRLLSSECLLPQPSRRRPRRSRGRGPAPR
                     ASCAGRPGSHGPRCTGCTTGSRCDPMGRVKVQGGGGSLVITQIGLQDAGYYQCVAENS
                     AGTACAAAPLAVVVREGLPSAPTRVTATPLSSSSVLVAWERPELHSEQIIGFSLHYQK
                     ARGVDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQLGASRTSSPALVHTLDDV
                     PSAAPQLTLSSPNPSDIRVAWLPLPSSLSNGQVLKYKIEYGLGKEEDQVFSTEVPGNE
                     TQLTLNSLQPNKVYRVRISAGTGAGYGVPSQWMQHRTPGVHNQSHVPFAPAELKVRAK
                     MESLVVSWQPPPHPTQISGYKLYWREVGTEEEADGDRPPGGRGDQAWDVGPVRLKKKV
                     KQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPTPDLPIQRGPPLPPAHVH
                     AESNSSTSIWLRWKKPDFTTVKIVNYTVRFGPWGLRNASLVTYYTSSGEDILIGGLKP
                     FTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTE
                     PNGEIVEYLILYSNNHTQPEHQWTLLTTEGNIFSAEVHGLESDTRYFFKMGARTEVGP
                     GPFSRLQDVITLQKTFSDSLDVHAVTGIIVGVCLGLLCLLACMCAGLRRSSHREALPG
                     LSSSGTPGNPALYTRARLGPPSVPAAHELESLVHPRPQDWSPPPSDVEDKAEVHSLMG
                     GSVSDCRGHSKRKISWAQAGGPNWAGSWAGCELPQGSGPRPALTRALLPPAGTGQTLL
                     LQALVYDAIKSNGRKKPSPACRNQVEAEVIVHSDFGASKGCPDLHLQDLEPEEPLTAE
                     TLPSTSGAVDLSQGADWLGRELGGCQPTTSGPERLTCLPEAASASCSCSDLQPSTAIE
                     EAPGKSCQPKALCPLTVSPSLPRAPVSSAQVP"
BASE COUNT     1362 a   1770 c   1669 g   1500 t
ORIGIN

  Query Match             92.8%;  Score 2595.2;  DB 10;  Length 6301;
  Best Local Similarity   94.5%;  Pred. No. 0;
  Matches 2783;  Conservative    0;  Mismatches   13;  Indels  150;  Gaps    3;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         42 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 101

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 161

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 221

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 281

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 341

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 401

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 461

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 521

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
              ||||||||||| ||||||||||||||||||||||  |  |||||||||||||||||||||
Db        522 GTGCCTGAGGAGTCCCGGCTCATCACTCTTCCCAATGGCGTCCTCCAGATCCTAGATGTC 581

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 641

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 701

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        702 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAGTGTAGTG 761

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAG------ 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     762 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGGGGAT 821

Qy     775 ------------------------------------------------------------ 774

Db     822 CAGAGTAAGAAGGAACTGGGGATTCAAACCAGCGGCTGTGATGCAGCTTTGCAGGGATGT 881

Qy     775 ------------------------------------------------------------ 774

Db     882 AGTGGGAACCTCGTAGTTGGAGAGGGTGTAGAGGGGACAGTTCCAGCGTCCCCCTCCAGC 941

Qy     775 --------------------GATGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCC 814
                                   ||||||||||||||||||||||||||||||||||||||||
Db     942 CTGCTCCTTCTCTGTCACCAGATGGAAAGCCTATCTCCACGGATGTCATCGTTCTGGGCC 1001

Qy     815 GGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAG 874
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1002 GGACCAATCTACTCATCGCCAGCGCGCAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAG 1061

Qy     875 CCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTG 934
           |||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||||
Db    1062 CCAACAAGCCCCGCACGCGTGATTTCGCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTG 1121

Qy     935 CCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCT 994
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1122 CCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCT 1181

Qy     995 TCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCC 1054
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1182 TCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCC 1241

Qy    1055 CGTTGCGACCCAATGGG-CGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACT 1113
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1242 CGTTGCGACCCAATGGGCCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACT 1301

Qy    1114 CAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACT 1173
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1302 CAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACT 1361

Qy    1174 GCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACT 1233
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1362 GCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACT 1421

Qy    1234 CGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAG 1293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1422 CGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAG 1481

Qy    1294 TTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGAC 1353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1482 TTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGAC 1541

Qy    1354 AATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTG 1413
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1542 AATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGGGACCTG 1601

Qy    1414 GAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGA 1473
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1602 GAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGA 1661

Qy    1474 ACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTT 1533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1662 ACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTT 1721

Qy    1534 ACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGC 1593
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1722 ACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGC 1781

Qy    1594 CTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAG---GAAGAT 1650
           |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db    1782 CTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAAGAAGAT 1841

Qy    1651 CAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAG 1710
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1842 CAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTTCAG 1901

Qy    1711 CCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCT 1770
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1902 CCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTCCCT 1961

Qy    1771 TCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCC 1830
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1962 TCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTTGCC 2021

Qy    1831 CCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCC 1890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2022 CCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCC 2081

Qy    1891 CCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGGAACAGAGGAG 1950
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    2082 CCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGAGAGAGGTGGGAACAGAGGAG 2141

Qy    1951 GAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCC 2010
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2142 GAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGGCCC 2201

Qy    2011 GTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCCG 2070
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    2202 GTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGGCTG 2261

Qy    2071 TACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAG 2130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2262 TACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGGAAG 2321

Qy    2131 GGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCT 2190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2322 GGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTGCCT 2381

Qy    2191 CCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAG 2250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2382 CCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAGAAG 2441

Qy    2251 CCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTC 2310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2442 CCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGGCTC 2501

Qy    2311 AGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGC 2370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2502 AGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGCGGC 2561

Qy    2371 CTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGG 2430
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2562 CTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGATGGG 2621

Qy    2431 CCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCT 2490
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2622 CCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCTTCT 2681

Qy    2491 GACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACG 2550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2682 GACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCCACG 2741

Qy    2551 GAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCC 2610
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2742 GAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAGCCC 2801

Qy    2611 GAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGC 2670
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2802 GAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCATGGC 2861

Qy    2671 CTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGG 2730
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2862 CTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGG 2921

Qy    2731 CCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGAT 2790
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    2922 CCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAAAGACATTCTCAGACTCCTTGGAT 2981
```

```
Qy       2791 GTGCAC 2796
              ||||||
Db       2982 GTGCAC 2987


RESULT 8
AB052622
LOCUS       AB052622                6485 bp    mRNA    linear   PRI 16-MAR-2002
DEFINITION  Homo sapiens hDDM36 mRNA, complete cds.
ACCESSION   AB052622
VERSION     AB052622.1  GI:19570397
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Murakami,H., Nakamata,T., Nakayama,T., Yamamoto,H., Hosaka,T.,
            Aoyama,T., Nagayama,S., Oka,M., Kiyono,T., Sasaki,M.S., Nakamura,T.
            and Toguchida,J.
  TITLE     Up-regulation of a ras effector and down-regulation of a cell
            adhesion molecule are associated with transformation of osteoblasts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6485)
  AUTHORS   Toguchida,J., Nakamata,T., Murakami,H., Nakayama,T. and Nakamura,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Junya Toguchida, Kyoto University,
            Institute for Frontier Medical Sciences; 53 Kawahara-cho, Shogoin,
            Sakyo-ku, Kyoto, Kyoto 606-8507, Japan
            (E-mail:togjun@frontier.kyoto-u.ac.jp, Tel:81-75-751-4134,
            Fax:81-75-751-4144)
FEATURES             Location/Qualifiers
     source          1..6485
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="hDDM36"
                     /cell_line="NHOst"
                     /cell_type="osteoblast"
                     /note="similar to mouse DDM36 and Nope
                     (NM_020043)-vector:phDDM36"
     gene            1..6485
                     /gene="hDDM36"
     CDS             211..3963
                     /gene="hDDM36"
                     /note="similar to mouse DDM36 and Nope (NP_064427)"
                     /codon_start=1
                     /evidence=experimental
                     /product="hDDM36"
                     /protein_id="BAB86306.1"
                     /db_xref="GI:19570398"
                     /translation="MARGDAGRGRGLLALTFCLLAARGELLLPQETTVELSCGVGPLQ
                     VILGPEQAAVLNCSLGAAAAGPPTRVTWSKDGDTLLEHDHLHLLPNGSLWLSQPLAPN
                     GSDESVPEAVGVIEGNYSCLAHGPLGVLASQTAVVKLATLADFSLHPESQTVEENGTA
                     RFECHIEGLPAPIITWEKDQVTLPEEPRLIVLPNGVLQILDVQESDAGPYRCVATNSA
                     RQHFSQEALLSVAHRGSLASTRGQDVVIVAAPENTTVVSGQSVVMECVASADPTPFVS
                     WVRQDGKPISTDVIVLGRTNLLIANAQPWHSGVYVCRANKPRTRDFATAAAELRVLAA
                     PAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNGAPLRPNGRVKVQGGGGSLVI
                     TQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPSAPTRVTATPLSSSAVLVAWE
                     RPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTELQVRDLEPNTDYEFYVVAYSQ
                     LGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVAWLPLPPSLSNGQVVKYKIEY
                     GLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAGTAAGFGAPSQWMHHRTPSMH
                     NQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYKLYWREVGAEEEANGDRLPGG
                     RGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAFNKHEDGYAAVWKGKTEKAPA
                     PDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTVKIVNYTVRFSPWGLRNASLV
                     TYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGSVVERSTLPDRPSTPPSDLRL
                     SPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEHQWTLLTTQGNIFSAEVHGLE
                     SDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLDMHSVTGIIVGVCLGLLCLLA
                     CMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPPSPPAAHELESLVHPHPQDWS
                     PPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSGLSWAGSWAGCELPQAGPRPA
                     LTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACRNQVEAEVIVHSDFSASNGNP
                     DLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDRELGGCELAAPGPDRLTCLPEAA
                     SASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLPRSPVSSSA"
BASE COUNT     1293 a   1926 c   1772 g   1494 t
ORIGIN

  Query Match             77.2%;  Score 2159.2;  DB 9;  Length 6485;
  Best Local Similarity   86.0%;  Pred. No. 0;
  Matches 2405;  Conservative    0;  Mismatches  388;  Indels    3;  Gaps    1;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
Db        280 GGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACTG 339

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
Db        340 CAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGCC 399

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
Db        400 GCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGAC 459

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
Db        460 CACTTACACCTGCTGCCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGGC 519

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
Db        520 AG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCTA 576

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
Db        577 GCCCACGGCCCCCTCGGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCACACTC 636

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
Db        637 GCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTTT 696

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
Db        697 GAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGACA 756

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
Db        757 TTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGTT 816

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
Db        817 CAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTTC 876

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
Db        877 AGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGGTCCCTGGCGTCCACCAGGGGGCAG 936

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
Db        937 GACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGTG 996

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
Db        997 ATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAAGACGGG 1056

Qy        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
Db       1057 AAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCG 1116

Qy        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
Db       1117 CAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTC 1176

Qy        901 GCCACTGCGGCTGCTGAGCTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAG 960
Db       1177 GCCACTGCAGCCGCTGAGCTCCGTGTGCTGGCGGCTCCCGCCATCACTCAGGCGCCCGAG 1236

Qy        961 GCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCA 1020
```

```
Db    1237 GCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGAGCCG 1296

Qy    1021 CGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAG 1080

Db    1297 CGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAG 1356

Qy    1081 GTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTAC 1140

Db    1357 GTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGCTAC 1416

Qy    1141 TACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTA 1200

Db    1417 TACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTG 1476

Qy    1201 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGC 1260

Db    1477 GTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGCAGC 1536

Qy    1261 TCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTC 1320

Db    1537 TCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGCTTC 1596

Qy    1321 TCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAAAC 1380

Db    1597 TCTCTCCACTACCAGAAGGCACGGGGCATGGACAATGTGGAATACCAGTTTGCAGTGAAC 1656

Qy    1381 AATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTCTAC 1440

Db    1657 AACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTCTAC 1716

Qy    1441 GTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCATACA 1500

Db    1717 GTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCACACA 1776

Qy    1501 CTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCGGAC 1560

Db    1777 CTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCGGAC 1836

Qy    1561 ATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAGTAC 1620

Db    1837 ATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAGTAC 1896

Qy    1621 AAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGAAAT 1680

Db    1897 AAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGAAAT 1956

Qy    1681 GAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATTTCA 1740

Db    1957 GAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATTTCG 2016

Qy    1741 GCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCTGGT 1800

Db    2017 GCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCCAGT 2076

Qy    1801 GTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAGATG 1860

Db    2077 ATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATG 2136

Qy    1861 GAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATACAAA 1920

Db    2137 GAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTACAAA 2196

Qy    1921 CTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGGGGT 1980

Db    2197 CTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGGGGC 2256

Qy    1981 CGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTAT 2040

Db    2257 CGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTAT 2316

Qy    2041 GAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAACAAA 2100

Db    2317 GAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAACAAA 2376

Qy    2101 CACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCAGAC 2160

Db    2377 CATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCAGAC 2436

Qy    2161 CTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAACAGC 2220

Db    2437 ATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAACAGC 2496

Qy    2221 TCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTCAAC 2280

Db    2497 TCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTCAAC 2556

Qy    2281 TACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTATACC 2340

Db    2557 TACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTACACC 2616

Qy    2341 AGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTTGCG 2400

Db    2617 AGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTTGCA 2676

Qy    2401 GTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCCACC 2460

Db    2677 GTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACC 2736

Qy    2461 CTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCATCC 2520

Db    2737 CTGCCTGACCGGCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCGTCC 2796

Qy    2521 ACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTATCTA 2580

Db    2797 ACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTATCTG 2856

Qy    2581 ATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCACAGAG 2640

Db    2857 ATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACGCAG 2916

Qy    2641 GGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTATTTCTTCAAG 2700

Db    2917 GGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTACTTCTTCAAG 2976

Qy    2701 ATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGATGTGATTACT 2760

Db    2977 ATGGGGGCGCGCACAGAGGTGGACCTGGGCCTTTCTCCCGCCTGCAGGATGTGATCACG 3036

Qy    2761 CTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796

Db    3037 CTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCAC 3072
```

```
RESULT 9
AX418406
LOCUS       AX418406                3741 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0206329.
ACCESSION   AX418406
VERSION     AX418406.1  GI:21523336
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Rastelli,L., Shimkets,R.A., Zerhusen,B., Malyankar,U.M. and
            Padigaru,M.
  TITLE     Human polynucleotides and polypeptides encoded thereby
  JOURNAL   Patent: WO 0206329-A 1 24-JAN-2002;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1..3741
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
                     /note="NOVX nucleic acid with homology to NOPE-like
```

```
                         protein"
BASE COUNT      708 a   1209 c   1150 g    673 t      1 others
ORIGIN

  Query Match             74.8%;  Score 2090.8;  DB 6;  Length 3741;
  Best Local Similarity   85.3%;  Pred. No. 0;
  Matches 2394;  Conservative    0;  Mismatches  393;  Indels   18;  Gaps    5;

Qy          1 GGGGAGCTGCCATTGCCCCAGGAGACAACTGTCAAGCTGAGCTGTGATGAGGGACCCCTG 60
Db         70 GGGGAGCTGCTGTTGCCCCAGGAGACGACTGTGGAGCTGAGCTGTGGAGTGGGGCCACTG 129

Qy         61 CAAGTGATCCTGGGCCCTGAGCAGGCTGTGGTGCTGGACTGCACTTTGGGGGCTACAGCT 120
Db        130 CAAGTGATCCTGGGCCCAGAGCAGGCTGCAGTGCTAAACTGTAGCCTGGGGGCTGCTGCC 189

Qy        121 GCTGGGCCTCCGACCAGGGTGACATGGAGCAAGGATGGAGACACTGTACTAGAGCATGAG 180
Db        190 GCTGGACCCCCCACCAGGGTGACCTGGAGCAAGGATGGGGACACCCTGCTGGAGCACGAC 249

Qy        181 AACCTGCACCTGCTACCCAATGGCTCCCTGTGGCTGTCCTCACCCCTAGAGCAAGAAGAC 240
Db        250 CACTTACACCTGCTGGCCAATGGTTCCCTGTGGCTGTCCCAGCCACTAGCACCCAATGGC 309

Qy        241 AGCGATGATGAGGAAGCTCTTAGGATCTGGAAGGTCACTGAGGGCAGCTATTCCTGTCTG 300
Db        310 AG---TGACGAGTCAGTCCCTGAGGCTGTGGGGGTCATTGAAGGCAACTATTCGTGCCTA 366

Qy        301 GCCCACAGCCCGCTAGGAGTGGTGGCCAGCCAGGTTGCTGTGGTCAAGCTTGCCACACTC 360
Db        367 GCCCACGGNCCCCCTCGAGTGCTGGCCAGCCAGACTGCTGTCGTCAAGCTTGCCAGTCTC 426

Qy        361 GAAGACTTCTCTCTGCACCCCGAGTCCCAGATTGTGGAGGAGAACGGGACAGCACGCTTT 420
Db        427 GCAGACTTCTCTCTGCACCCGGAGTCTCAGACGGTGGAGGAGAACGGGACAGCTCGCTTT 486

Qy        421 GAATGCCACACCAAGGGCCTTCCAGCCCCCATCATTACTTGGGAAAAGGACCAGGTGACC 480
Db        487 GAGTGCCACATTGAAGGGCTGCCAGCTCCCATCATTACTTGGGAGAAGGACCAGGTGACA 546

Qy        481 GTGCCTGAGGAGCCCCGGCTCATCACTCTTCCCAAGTGGCTCCTCCAGATCCTAGATGTC 540
Db        547 TTGCCTGAGGAGCCTCGGCTCATCGTGCTTCCCAACGGCGTCCTTCAGATCCTGGATGTT 606

Qy        541 CAGGACAGTGATGCAGGCTCCTACCGCTGCGTGGCCACCAATTCAGCCCGCCAACGATTC 600
Db        607 CAGGAGAGTGATGCAGGCCCCTACCGCTGCGTGGCCACCAACTCAGCTCGCCAGCACTTC 666

Qy        601 AGCCAGGAGGCCTCGCTCACTGTGGCCCTCAGAGGGTCTTTGGAGGCTACCAGGGGGCAG 660
Db        667 AGCCAGGAGGCCCTACTCAGTGTGGCCCACAGAGGTTCCCTGGCGTCCACCAGGGGGCAG 726

Qy        661 GATGTGGTCATTGTGGCAGCCCCAGAGAACACCACGGTAGTGTCTGGACAGAATGTAGTG 720
Db        727 GACGTGGTCATTGTGGCAGCCCCAGAGAACACCACAGTGGTGTCTGGCCAGAGTGTGGTG 786

Qy        721 ATGGAGTGCGTGGCCTCTGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGACAGGATGGA 780
Db        787 ATGGAATGTGTGGCCTCAGCTGACCCCACCCCTTTTGTGTCCTGGGTCCGA---GACGGG 843

Qy        781 AAGCCTATCTCCACGGATGTCATCGTTCTGGGCCGGACCAATCTACTCATCGCCAGCGCG 840
Db        844 AAGCCCATCTCCACAGATGTCATCGTCCTGGGCCGCACCAACCTACTAATTGCCAACGCG 903

Qy        841 CAGCCTCGGCACTCTGGAGTCTATGTCTGCCGAGCCAACAAGCCCCTCACGCGTGACTTC 900
Db        904 CAGCCCTGGCACTCCGGCGTCTATGTCTGCCGCGCCAACAAGCCCCGCACGCGCGACTTC 963

Qy        901 GCCACTGCGGCTGCTGAGCTCCGAG---TGCTTGCTGCCCCAGCCATCTCGCAGGCACCC 957
Db        964 GCCACTGCAGCCGCTGAGCTCCGTGTGCTGCTAGCGGCTCCCGCCATCACTCAGGCGCCC 1023

Qy        958 GAGGCGCTCTCGCGGACGCGGGCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAG 1017
Db       1024 GAGGCGCTGTCGCGGACGCGGGCGAGCACAGCGCGCTTCGTGTGCCGCGCGTCGGGGGAG 1083

Qy       1018 CCACGGCCCGCGCTGCACTGGCTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTC 1077
Db       1084 CCGCGGCCAGCGCTGCGCTGGCTGCACAACGGGGCGCCGCTGCGGCCCAACGGGCGCGTC 1143

Qy       1078 AAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGC 1137
Db       1144 AAGGTCCAGGGCGGCGGTGGCAGCCTGGTCATCACACAGATCGGCCTGCAGGACGCCGGC 1203

Qy       1138 TACTACCAGTGCGTAGCAGAAAACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCG 1197
Db       1204 TACTACCAGTGCGTGGCTGAGAACAGCGCGGGAATGGCGTGCGCTGCCCGCGTCGCTGGCC 1263

Qy       1198 GTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGC 1257
Db       1264 GTGGTGGTGCGCGAGGGGCTGCCCAGCGCCCCCACGCGGGTCACTGCTACGCCACTGAGC 1323

Qy       1258 AGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGC 1317
Db       1324 AGCTCCGCTGTGTTGGTGGCCTGGGAGCGGCCCGAGATGCACAGCGAGCAGATCATCGGC 1383

Qy       1318 TTCTCTCTTCACTACCAAAAGGCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTA 1377
Db       1384 TTCTCTCTCCACTACCAGAAGGCACGGGGTATGGACAATGTGGAATACCAGTTTGCAGTG 1443

Qy       1378 AACAATGACACCACAGAGCTGCAGGTTCGGGACCTGGAACCCAACACGGATTATGAGTTC 1437
Db       1444 AACAACGACACCACAGAACTACAGGTTCGGGACCTGGAACCCAACACAGATTATGAGTTC 1503

Qy       1438 TACGTGGTGGCCTACTCCCAGCTGGGGGCCAGCCGAACCTCCAGCCCAGCCCTGGTGCAT 1497
Db       1504 TACGTGGTGGCCTACTCCCAGCTGGGAGCCAGCCGCACCTCCACCCCAGCACTGGTGCAC 1563

Qy       1498 ACACTGGACGATGTCCCCAGCGCAGCACCCCAGCTTACCTTGTCCAGCCCCAACCCCTCG 1557
Db       1564 ACACTGGATGATGTCCCCAGTGCAGCACCCCAGCTCTCCCTGTCCAGCCCCAACCCTTCG 1623

Qy       1558 GACATCAGGGTGGCATGGCTGCCCCTGCCCTCCAGCCTGAGCAATGGACAGGTGCTGAAG 1617
Db       1624 GACATCAGGGTGGCGTGGCTGCCCCTGCCCCCCAGCCTGAGCAATGGGCAGGTGGTGAAG 1683

Qy       1618 TACAAGATAGAGTACGGTTTGGGGAAGGAAGATCAGGTTTTCTCCACCGAGGTGCCTGGA 1677
Db       1684 TACAAGATAGAATACGGTTTGGGAAAGGAAGATCAGATTTTCTCTACTGAGGTGCGAGGA 1743

Qy       1678 AATGAGACACAACTTACGTTAAACTCACTTCAGCCAAACAAAGTGTACCGAGTCCGGATT 1737
Db       1744 AATGAGACACAGCTTATGCTGAACTCGCTTCAGCCAAACAAGGTGTATCGAGTACGGATT 1803

Qy       1738 TCAGCTGGCACTGGCGCTGGCTATGGAGTCCCTTCTCAGTGGATGCAGCACAGGACACCT 1797
Db       1804 TCGGCTGGTACAGCAGCCGGCTTCGGGGCCCCCTCCCAGTGGATGCATCACAGGACGCCC 1863

Qy       1798 GGTGTGCACAACCAGAGCCATGTTCCCTTTGCCCCTGCAGAATTGAAGGTGAGGGCAAAG 1857
Db       1864 AGTATGCACAACCAGAGCCATGTCCCTTTTGCCCCTGCAGAGTTGAAGGTGCAGGCAAAG 1923

Qy       1858 ATGGAGTCCCTGGTGGTGTCATGGCAGCCGCCCCCTCACCCCACCCAGATCTCTGGATAC 1917
Db       1924 ATGGAGTCCCTGGTCGTGTCATGGCAGCCACCCCCTCACCCCACCCAGATCTCTGGCTAC 1983

Qy       1918 AAACTCTACTGGGGAGAGGTGGGAACAGAGGAGGAGGCAGATGGTGACCGCCCCCCAGGG 1977
Db       1984 AAACTATATTGGCGGGAGGTGGGGGCTGAGGAGGAGGCCAATGGCGATCGCCTGCCAGGG 2043

Qy       1978 GGTCGTGGAGATCAAGCTTGGGACGTCGGGCCCGTGCGGCTGAAGAAGAAAGTGAAGCAG 2037
Db       2044 GGCCGTGGAGACCAGGCTTGGGATGTGGGGCCTGTCCGGCTCAAGAAGAAAGTGAAGCAG 2103

Qy       2038 TATGAACTGACCCAGTTAGTCCCTGGCAGGCCGTACGAGGTGAAGCTCGTAGCTTTCAAC 2097
```

```
Db     2104 TATGAGCTGACCCAGCTAGTCCCTGGCCGGCTGTACGAGGTGAAGCTCGTGGCTTTCAAC 2163

Qy     2098 AAACACGAGGACGGCTACGCTGCTGTGTGGAAGGGCAAGACGGAGAAGGCGCCCACGCCA 2157

Db     2164 AAACATGAGGATGGCTATGCAGCAGTGTGGAAGGGCAAGACGGAGAAGGCGCCGGCACCA 2223

Qy     2158 GACCTGCCTATCCAGAGGGGGCCACCGCTGCCTCCTGCCCATGTCCACGCAGAGTCAAAC 2217

Db     2224 GACATGCCTATCCAGAGGGGACCACCCCTGCCTCCAGCCCACGTCCATGCGGAATCAAAC 2283

Qy     2218 AGCTCCACTTCCATTTGGCTTCGGTGGAAGAAGCCAGACTTTACCACTGTCAAGATTGTC 2277

Db     2284 AGCTCCACATCCATCTGGCTTCGGTGGAAAAAGCCAGATTTCACCACAGTCAAGATTGTC 2343

Qy     2278 AACTACACTGTACGCTTCGGCCCCTGGGGGCTCAGGAATGCTTCCCTGGTCACCTACTAT 2337

Db     2344 AACTACACTGTGCGCTTCAGCCCCTGGGGGCTCAGGAATGCCTCCCTGGTCACCTATTA- 2402

Qy     2338 ACCAGCTCTGGAGAAGACATTCTCATTGGCGGCCTGAAACCATTTACCAAGTACGAGTTT 2397

Db     2403 --CAGTTCTGGAGAAGACATCCTCATTGGCGGCTTGAAGCCATTCACCAAATACGAGTTT 2460

Qy     2398 GCGGTACAGTCCCACGGAGTGGATATGGATGGGCCCTTTGGCTCCGTCGTAGAACGCTCC 2457

Db     2461 GCAGTGCAGTCTCACGGCGTGGACATGGATGGGCCTTTCGGCTCTGTGGTGGAGCGCTCC 2520

Qy     2458 ACCCTGCCTGACCGGCCTTCAACACCTCCTTCTGACCTGCGCCTGAGCCCCCTGACACCA 2517

Db     2521 ACCCTGCCTGACCGTCCCTCCACACCCCCATCCGACCTGCGACTGAGCCCCCTGACACCG 2580

Qy     2518 TCCACCGTTCGGTTACACTGGTGTCCCCCCACGGAGCCCAATGGTGAGATTGTGGAGTAT 2577

Db     2581 TCCACGGTTCGGCTGCACTGGTGCCCCCCCACAGAGCCCAACGGGGAGATCGTGGAGTAT 2640

Qy     2578 CTAATTCTCTACAGCAACAACCACACCCAGCCCGAACACCAGTGGACACTGCTCACCAC- 2636

Db     2641 CTGATCCTGTACAGCAGCAACCACACGCAGCCTGAGCACCAGTGGACCTTGCTCACCACG 2700

Qy     2637 -----AGAGGGAAACATCTTCAGTGCAGAGGTCCATGGCCTAGAGAGTGACACTCGGTAT 2691

Db     2701 CAGGGTGAGGGAAACATCTTCAGTGCTGAGGTCCATGGCCTGGAGAGCGACACTCGGTAC 2760

Qy     2692 TTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCTGGGCCCTTTTCCCGCTTGCAGGAT 2751

Db     2761 TTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCTGGGCCTTTCTCCCGCCTGCAGGAT 2820

Qy     2752 GTGATTACTCTGCAAGAGACATTCTCAGACTCCTTGGATGTGCAC 2796

Db     2821 GTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTGGACATGCAC 2865
```

```
RESULT 10
AB046848
LOCUS       AB046848                5549 bp    mRNA    linear   PRI 22-FEB-2001
DEFINITION  Homo sapiens mRNA for KIAA1628 protein, partial cds.
ACCESSION   AB046848
VERSION     AB046848.1  GI:10047332
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Kikuno,R., Nakayama,M., Hirosawa,M. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XVIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 7 (4), 273-281 (2000)
  MEDLINE   20450683
   PUBMED   10997877
REFERENCE   2  (bases 1 to 5549)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..5549
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="fh14247"
                     /note="vector:pBluescriptII SK plus"
     gene            1..5549
                     /gene="KIAA1628"
     CDS             <81..3023
                     /gene="KIAA1628"
                     /note="Start codon is not identified. fh14247 cDNA clone
                     for KIAA1628 has a 40-bp different sequence replacing a
                     108-bp sequence at the position between 737 and 846, and
                     an 88-bp insertion after the position 983 of the sequence
                     of KIAA1628, respectively."
                     /codon_start=1
                     /product="KIAA1628 protein"
                     /protein_id="BAB13454.1"
                     /db_xref="GI:10047333"
                     /translation="MGRRKGRSALSRSSLERVNQGEVKVIGRGAREGGGGQAWGTGGG
                     GRDLSCGIPRSASPLAPLAAPAITQAPEALSRTRASTARFVCRASGEPRPALRWLHNG
                     APLRPNGRVKVQGGGGSLVITQIGLQDAGYYQCVAENSAGMACAAASLAVVVREGLPS
                     APTRVTATPLSSSAVLVAWERPEMHSEQIIGFSLHYQKARGMDNVEYQFAVNNDTTEL
                     QVRDLEPNTDYEFYVVAYSQLGASRTSTPALVHTLDDVPSAAPQLSLSSPNPSDIRVA
                     WLPLPPSLSNGQVVKYKIEYGLGKEDQIFSTEVRGNETQLMLNSLQPNKVYRVRISAG
                     TAAGFGAPSQWMHHRTPSMHNQSHVPFAPAELKVQAKMESLVVSWQPPPHPTQISGYK
                     LYWREVGAEEEANGDRLPGGRGDQAWDVGPVRLKKKVKQYELTQLVPGRLYEVKLVAF
                     NKHEDGYAAVWKGKTEKAPAPDMPIQRGPPLPPAHVHAESNSSTSIWLRWKKPDFTTV
                     KIVNYTVRFSPWGLRNASLVTYYTSSGEDILIGGLKPFTKYEFAVQSHGVDMDGPFGS
                     VVERSTLPDRPSTPPSDLRLSPLTPSTVRLHWCPPTEPNGEIVEYLILYSSNHTQPEH
                     QWTLLTTQGNIFSAEVHGLESDTRYFFKMGARTEVGPGPFSRLQDVITLQEKLSDSLD
                     MHSVTGIIVGVCLGLLCLLACMCAGLRRSPHRESLPGLSSTATPGNPALYSRARLGPP
                     SPPAAHELESLVHPHPQDWSPPPSDVEDRAEVHSLMGGGVSEGRSHSKRKISWAQPSG
                     LSWAGSWAGCELPQAGPRPALTRALLPPAGTGQTLLLQALVYDAIKGNGRKKSPPACR
                     NQVEAEVIVHSDFSASNGNPDLHLQDLEPEDPLPPEAPDLISGVGDPGQGAAWLDREL
                     GGCELAAPGPDRLTCLPEAASASCSYPDLQPGEVLEETPGDSCQLKSPCPLGASPGLP
                     RSPVSSSA"
BASE COUNT     1157 a   1546 c   1496 g   1350 t
ORIGIN

  Query Match             53.3%;  Score 1491.4;  DB 9;  Length 5549;
  Best Local Similarity   87.4%;  Pred. No. 3.3e-298;
  Matches 1633;  Conservative    0;  Mismatches  236;  Indels    0;  Gaps    0;
```

```
Qy      928 CTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACC 987

Db      264 CTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCACA 323

Qy      988 GCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGAC 1047

Db      324 GCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAAC 383

Qy     1048 GGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTC 1107

Db      384 GGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGTC 443

Qy     1108 ATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCG 1167

Db      444 ATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGCG 503

Qy     1168 GGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCC 1227

Db      504 GGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGCC 563

Qy     1228 CCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGG 1287

Db      564 CCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCGG 623
```

```
Qy   1288 CCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGA 1347
Db    624 CCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGGC 683

Qy   1348 GTGGACAATGTGGAGTACCAGTTTGCAGTAAACAATGACACCACAGAGCTGCAGGTTCGG 1407
Db    684 ATGGACAATGTGGAATACCAGTTTGCAGTGAACAACGACACCACAGAACTACAGGTTCGG 743

Qy   1408 GACCTGGAACCCAACACGGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGGGCC 1467
Db    744 GACCTGGAACCCAACACAGATTATGAGTTCTACGTGGTGGCCTACTCCCAGCTGGGAGCC 803

Qy   1468 AGCCGAACCTCCAGCCCAGCCCTGGTGCATACACTGGACGATGTCCCCAGCGCAGCACCC 1527
Db    804 AGCCGCACCTCCACCCCAGCACTGGTGCACACACTGGATGATGTCCCCAGTGCAGCACCC 863

Qy   1528 CAGCTTACCTTGTCCAGCCCCAACCCCTCGGACATCAGGGTGGCATGGCTGCCCCTGCCC 1587
Db    864 CAGCTCTCCCTGTCCAGCCCCAACCCTTCGGACATCAGGGTGGCGTGGCTGCCCCTGCCC 923

Qy   1588 TCCAGCCTGAGCAATGGACAGGTGCTGAAGTACAAGATAGAGTACGGTTTGGGGAAGGAA 1647
Db    924 CCCAGCCTGAGCAATGGGCAGGTGGTGAAGTACAAGATAGAATACGGTTTGGGAAAGGAA 983

Qy   1648 GATCAGGTTTTCTCCACCGAGGTGCCTGGAAATGAGACACAACTTACGTTAAACTCACTT 1707
Db    984 GATCAGATTTTCTCTACTGAGGTGCGAGGAAATGAGACACAGCTTATGCTGAACTCGCTT 1043

Qy   1708 CAGCCAAACAAAGTGTACCGAGTCCGGATTTCAGCTGGCACTGGCGCTGGCTATGGAGTC 1767
Db   1044 CAGCCAAACAAGGTGTATCGAGTACGGATTTCGGCTGGTACAGCAGCCGGCTTCGGGGCC 1103

Qy   1768 CCTTCTCAGTGGATGCAGCACAGGACACCTGGTGTGCACAACCAGAGCCATGTTCCCTTT 1827
Db   1104 CCCTCCCAGTGGATGCATCACAGGACGCCCAGTATGCACAACCAGAGCCATGTCCCTTTT 1163

Qy   1828 GCCCCTGCAGAATTGAAGGTGAGGGCAAAGATGGAGTCCCTGGTGGTGTCATGGCAGCCG 1887
Db   1164 GCCCCTGCAGAGTTGAAGGTGCAGGCAAAGATGGAGTCCCTGGTCGTGTCATGGCAGCCA 1223

Qy   1888 CCCCCTCACCCCACCCAGATCTCTGGATACAAACTCTACTGGGGAGAGGTGGAACAGAG 1947
Db   1224 CCCCCTCACCCCACCCAGATCTCTGGCTACAAACTATATTGGCGGGAGGTGGGGGCTGAG 1283

Qy   1948 GAGGAGGCAGATGGTGACCGCCCCCCAGGGGGTCGTGGAGATCAAGCTTGGGACGTCGGG 2007
Db   1284 GAGGAGGCCAATGGCGATCGCCTGCCAGGGGGCCGTGGAGACCAGGCTTGGGATGTGGGG 1343

Qy   2008 CCCGTGCGGCTGAAGAAGAAAGTGAAGCAGTATGAACTGACCCAGTTAGTCCCTGGCAGG 2067
Db   1344 CCTGTCCGGCTCAAGAAGAAAGTGAAGCAGTATGAGCTGACCCAGCTAGTCCCTGGCCGG 1403

Qy   2068 CCGTACGAGGTGAAGCTCGTAGCTTTCAACAAACACGAGGACGGCTACGCTGCTGTGTGG 2127
Db   1404 CTGTACGAGGTGAAGCTCGTGGCTTTCAACAAACATGAGGATGGCTATGCAGCAGTGTGG 1463

Qy   2128 AAGGGCAAGACGGAGAAGGCGCCCACGCCAGACCTGCCTATCCAGAGGGGGCCACCGCTG 2187
Db   1464 AAGGGCAAGACGGAGAAGGCGCCGGCACCAGACATGCCTATCCAGAGGGGACCACCCCTG 1523

Qy   2188 CCTCCTGCCCATGTCCACGCAGAGTCAAACAGCTCCACTTCCATTTGGCTTCGGTGGAAG 2247
Db   1524 CCTCCAGCCCACGTCCATGCGGAATCAAACAGCTCCACATCCATCTGGCTTCGGTGGAAA 1583

Qy   2248 AAGCCAGACTTTACCACTGTCAAGATTGTCAACTACACTGTACGCTTCGGCCCCTGGGGG 2307
Db   1584 AAGCCAGATTTCACCACAGTCAAGATTGTCAACTACACTGTGCGCTTCAGCCCCTGGGGG 1643

Qy   2308 CTCAGGAATGCTTCCCTGGTCACCTACTATACCAGCTCTGGAGAAGACATTCTCATTGGC 2367
Db   1644 CTCAGGAATGCCTCCCTGGTCACCTATTACACCAGTTCTGGAGAAGACATCCTCATTGGC 1703

Qy   2368 GGCCTGAAACCATTTACCAAGTACGAGTTTGCGGTACAGTCCCACGGAGTGGATATGGAT 2427
Db   1704 GGCTTGAAGCCATTCACCAAATACGAGTTTGCAGTGCAGTCTCACGGCGTGGACATGGAT 1763

Qy   2428 GGGCCCTTTGGCTCCGTCGTAGAACGCTCCACCCTGCCTGACCGGCCTTCAACACCTCCT 2487
Db   1764 GGGCCTTTCGGCTCTGTGGTGGAGCGCTCCACCCTGCCTGACCGGCCCTCCACACCCCCA 1823

Qy   2488 TCTGACCTGCGCCTGAGCCCCCTGACACCATCCACCGTTCGGTTACACTGGTGTCCCCCC 2547
Db   1824 TCCGACCTGCGACTGAGCCCCCTGACACCGTCCACGGTTCGGCTGCACTGGTGCCCCCCC 1883

Qy   2548 ACGGAGCCCAATGGTGAGATTGTGGAGTATCTAATTCTCTACAGCAACAACCACACCCAG 2607
Db   1884 ACAGAGCCCAACGGGGAGATCGTGGAGTATCTGATCCTGTACAGCAGCAACCACACGCAG 1943

Qy   2608 CCCGAACACCAGTGGACACTGCTCACCACAGAGGGAAACATCTTCAGTGCAGAGGTCCAT 2667
Db   1944 CCTGAGCACCAGTGGACCTTGCTCACCACGCAGGGAAACATCTTCAGTGCTGAGGTCCAT 2003

Qy   2668 GGCCTAGAGAGTGACACTCGGTATTTCTTCAAGATGGGAGCCCGCACAGAGGTGGGGCCT 2727
Db   2004 GGCCTGGAGAGCGACACTCGGTACTTCTTCAAGATGGGGGCGCGCACAGAGGTGGGACCT 2063

Qy   2728 GGGCCCTTTTCCCGCTTGCAGGATGTGATTACTCTGCAAGAGACATTCTCAGACTCCTTG 2787
Db   2064 GGGCCTTTCTCCCGCCTGCAGGATGTGATCACGCTCCAGGAGAAGCTGTCAGACTCGCTG 2123

Qy   2788 GATGTGCAC 2796
Db   2124 GACATGCAC 2132
```

RESULT 11
AC112162/c

LOCUS	AC112162 116580 bp DNA linear ROD 21-SEP-2002
DEFINITION	Mus Musculus chromosome 9 BAC clone MGS1-296M6 ES cell line, complete sequence.
ACCESSION	AC112162
VERSION	AC112162.8 GI:21327375
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 116580)
AUTHORS	Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J., Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E., Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
TITLE	High Throughput Mouse Sequencing
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 116580)
AUTHORS	Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J., Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E., Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
TITLE	Direct Submission
JOURNAL	Submitted (20-FEB-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA
REFERENCE	3 (bases 1 to 116580)
AUTHORS	Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J., Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E., Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
TITLE	Direct Submission
JOURNAL	Submitted (04-MAY-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA
REFERENCE	4 (bases 1 to 116580)
AUTHORS	Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J., Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E., Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUN-2002) Harvard Partners Center for Genetics and

```
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   5  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   6  (bases 1 to 116580)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Brown,W.A., Chiu,D., Decker,J.,
            Fusina,M., Haider,A., Keller,A., Perera,A., Shim,C., Thomas,E.,
            Zencheck,W., Xi,C., Juels,P. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On Jun 6, 2002 this sequence version replaced gi:21321767.
            ----------------Genome Center:
            Center:   Harvard Partners Genome Center
            Center Code:   HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact:  hpgc@mendel.mgh.harvard.edu

            CLONE LENGTH: This sequence represents the entire insert of this
            clone unless otherwise noted. If there are overlapping clones, the
            overlaps are noted in the beginning and end of the Features
            listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550).
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
              Genes and Regions of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
            and cDNA sequences in Unigene. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintain sequence continuity
            across the splice junctions. Sequences that are not identical
            matches are annotated as similar.

            SEQUENCING READ COVERAGE: Attempts are made to complete double
            strandedsequence for all regions.  All sequence is completed to a
            standard of coverage with a minimum of 3 reads with no ambiguities.
            If the sequence coverage for a region does not meet this standard,
            it is indicated in the annotation asLow Coverage.  Low coverage
            linkages are verified by PCR product size verification or
            verification of forward and reverse reads from clones which span
            the low coverage area.

            QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
            standards - estimated average error rate is less than 1 per 10,000
            bases using the Consed quality parameters.  Regions that do not
            meet this requirement are annotated as Low Quality.

            ----------------Summary Statistics
            Center project name: ALI
            Sequencing vector: pSMART; AF399742
            Chemistry: Dye-terminator Big Dye; 100%
            Assembly program: Phrap version 0.990319
            Contig length: 116580
            Fraction of Phrap value < 40: 0
            Error Rate in Consed: 0.00 per 10,000 bases
            Number of N's in consensus: 0

            ----------- Distribution of Quality < 40 Bases:

                   1000|
                    900|
                    800|
             #      700|
             bases  600|
                    500|
                    400|
                    300|
                    200|
                    100|
                      0|
                        -----------------------------------------
                          5    10   15   20   25   30   35   40
                                   Phrap Value Range
                   ----------------------.
FEATURES             Location/Qualifiers
     source          1. .116580
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="ES cell line"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /clone="MGS1-296M6"
     STS             235. .369
                     /standard_name="D9Mit316"
     repeat_region   258. .303
                     /rpt_family="(TG)n"
     STS             612. .1204
                     /standard_name="Nope"
     repeat_region   complement(1584. .1643)
                     /rpt_family="B4A"
     repeat_region   1710. .1748
                     /rpt_family="(TG)n"
     repeat_region   complement(3793. .3909)
                     /rpt_family="PB1D10"
     repeat_region   4042. .4191
                     /rpt_family="MIR"
     repeat_region   4247. .4435
                     /rpt_family="B2_Mm2"
     repeat_region   complement(4504. .4623)
                     /rpt_family="RSINE1"
     repeat_region   4720. .4831
                     /rpt_family="L2"
     repeat_region   complement(4840. .4879)
                     /rpt_family="MIR"
     repeat_region   4938. .5118
                     /rpt_family="B2_Mm2"
     repeat_region   6216. .6250
                     /rpt_family="AT_rich"
     repeat_region   complement(6251. .6381)
                     /rpt_family="RSINE1"
     repeat_region   complement(6421. .6566)
                     /rpt_family="B1_MM"
     repeat_region   6593. .6612
                     /rpt_family="(T)n"
     repeat_region   complement(6668. .6790)
                     /rpt_family="B1_MM"
     repeat_region   complement(6705. .6820)
                     /rpt_family="MT2B"
     repeat_region   complement(6883. .6954)
                     /rpt_family="ID_B1"
     repeat_region   complement(6982. .7058)
                     /rpt_family="ID4"
     repeat_region   8050. .8246
                     /rpt_family="B2_Mm2"
     repeat_region   complement(8445. .8719)
                     /rpt_family="B4"
     repeat_region   complement(9019. .9119)
                     /rpt_family="MIR"
     repeat_region   9265. .9371
                     /rpt_family="MIR"
     repeat_region   complement(10773. .10846)
                     /rpt_family="PB1D10"
     repeat_region   10859. .10956
                     /rpt_family="B4"
     repeat_region   complement(11043. .11141)
                     /rpt_family="PB1D7"
     repeat_region   11848. .11884
                     /rpt_family="(TC)n"
```

```
     repeat_region   11884..11922
                     /rpt_family="(TG)n"
     repeat_region   11930..11966
                     /rpt_family="(TC)n"
     repeat_region   complement(11967..11994)
                     /rpt_family="B4"
     repeat_region   complement(11995..12127)
                     /rpt_family="B1_MM"
     repeat_region   complement(12128..12143)
                     /rpt_family="B4"
     repeat_region   complement(12170..12252)
                     /rpt_family="RSINE1"
     repeat_region   12853..12993
                     /rpt_family="B1_MM"
     repeat_region   complement(13288..13367)
                     /rpt_family="MIR"

  Query Match            14.8%;  Score 415;  DB 10;  Length 116580;
  Best Local Similarity  97.7%;  Pred. No. 1e-75;
  Matches  421;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        931 GCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 990
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      10502 GCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 10443

Qy        991 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10442 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 10383

Qy       1051 ATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10382 ATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATC 10323

Qy       1111 ACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGA 1170
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      10322 ACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGA 10263

Qy       1171 ACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCG 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10262 ACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCG 10203

Qy       1231 ACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCT 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10202 ACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCT 10143

Qy       1291 GAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTG 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db      10142 GAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGTATG 10083

Qy       1351 GACAATGTGGA 1361
                 | | | ||
Db      10082 TCTACTCTTGA 10072

RESULT 12
AC110235/c
LOCUS       AC110235               158055 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-100M12, WORKING DRAFT SEQUENCE, 17 ordered
            pieces.
ACCESSION   AC110235
VERSION     AC110235.5  GI:20800205
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-100M12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 15, 2002 this sequence version replaced gi:20531900.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22383
                Center clone name: 100_M_12
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 153807 bases at least Q40
                Consensus quality: 155134 bases at least Q30
                Consensus quality: 155784 bases at least Q20
                Insert size: 218000; agarose-fp
                Insert size: 156455; sum-of-contigs
                Quality coverage: 7.1 in Q20 bases; agarose-fp
                Quality coverage: 10.0 in Q20 bases; sum-of-contigs
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1      6459: contig of 6459 bp in length
            *     6460      6559: gap of 100 bp
            *     6560      6667: contig of 108 bp in length
            *     6668      6767: gap of 100 bp
            *     6768      7863: contig of 1096 bp in length
            *     7864      7963: gap of 100 bp
            *     7964      9426: contig of 1463 bp in length
            *     9427      9526: gap of 100 bp
            *     9527     11011: contig of 1485 bp in length
            *    11012     11111: gap of 100 bp
            *    11112     12741: contig of 1630 bp in length
            *    12742     12841: gap of 100 bp
            *    12842     14789: contig of 1948 bp in length
            *    14790     14889: gap of 100 bp
            *    14890     17826: contig of 2937 bp in length
            *    17827     17926: gap of 100 bp
            *    17927     21244: contig of 3318 bp in length
            *    21245     21344: gap of 100 bp
            *    21345     27413: contig of 6069 bp in length
            *    27414     27513: gap of 100 bp
            *    27514     36381: contig of 8868 bp in length
            *    36382     36481: gap of 100 bp
            *    36482     45104: contig of 8623 bp in length
            *    45105     45204: gap of 100 bp
            *    45205     58369: contig of 13165 bp in length
            *    58370     58469: gap of 100 bp
            *    58470     85782: contig of 27313 bp in length
            *    85783     85882: gap of 100 bp
            *    85883    112706: contig of 26824 bp in length
            *   112707    112806: gap of 100 bp
            *   112807    153927: contig of 41121 bp in length
            *   153928    154027: gap of 100 bp
            *   154028    158055: contig of 4028 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158055
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-100M12"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .6459
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    6560. .6667
                     /note="assembly_fragment"
     misc_feature    6768. .7863
                     /note="assembly_fragment"
     misc_feature    7964. .9426
                     /note="assembly_fragment"
     misc_feature    9527. .11011
                     /note="assembly_fragment"
     misc_feature    11112. .12741
                     /note="assembly_fragment"
     misc_feature    12842. .14789
                     /note="assembly_fragment"
     misc_feature    14890. .17826
                     /note="assembly_fragment"
     misc_feature    17927. .21244
                     /note="assembly_fragment"
     misc_feature    21345. .27413
                     /note="assembly_fragment"
     misc_feature    27514. .36381
                     /note="assembly_fragment"
     misc_feature    36482. .45104
                     /note="assembly_fragment"
     misc_feature    45205. .58369
                     /note="assembly_fragment"
     misc_feature    58470. .85782
                     /note="assembly_fragment"
     misc_feature    85883. .112706
                     /note="assembly_fragment"
     misc_feature    112807. .153927
                     /note="assembly_fragment"
     misc_feature    154028. .158055
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    41469 a  36830 c  36521 g  41635 t   1600 others
ORIGIN

  Query Match            14.8%;  Score 415;  DB 2;  Length 158055;
  Best Local Similarity  97.7%;  Pred. No. 9.9e-76;
  Matches  421;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        931 GCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 990
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      85019 GCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTCTCGCGGACGCGGGCCAGCACCGCG 84960

Qy        991 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 1050
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84959 CGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGACGGG 84900

Qy       1051 ATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATC 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84899 ATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGTCATC 84840

Qy       1111 ACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGA 1170
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      84839 ACTCAGATCGGCCTTCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGCGGGA 84780

Qy       1171 ACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCG 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84779 ACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGCCCCG 84720

Qy       1231 ACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCT 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84719 ACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCGGCCT 84660

Qy       1291 GAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGAGTG 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db      84659 GAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGGTATG 84600

Qy       1351 GACAATGTGGA 1361
                 | | | ||
Db      84599 TCTACTCTTGA 84589


RESULT 13
AC106509/c
LOCUS       AC106509              224227 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-202I8, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC106509
VERSION     AC106509.4  GI:30580817
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 224227)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
```

```
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuohu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 224227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 224227)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:22857495.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
              . Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLBA
                Center clone name: CH230-202I8
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 211258 bases at least Q40
                Consensus quality: 214044 bases at least Q30
                Consensus quality: 215992 bases at least Q20
                Estimated insert size: 220399; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   221791: contig of 221791 bp in length
            *   221792   221891: gap of unknown length
            *   221892   222909: contig of 1018 bp in length
            *   222910   223009: gap of unknown length
            *   223010   224227: contig of 1218 bp in length.
FEATURES             Location/Qualifiers
     source          1..224227
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-202I8"
     misc_feature    1..1752
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3617..3835
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH334357"
     misc_feature    220120..221791
                     /note="wgs_contig"
BASE COUNT    58284 a  52300 c  50849 g  55959 t   6835 others
ORIGIN

  Query Match            13.5%;  Score 376.8;  DB 2;  Length 224227;
  Best Local Similarity  88.7%;  Pred. No. 7.5e-68;
  Matches  408;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         919 CTCCGAGTGCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGG 978
               ||||     |  ||||||||||||||||||||||| ||||||||||| ||||||||  ||
Db      109894 CTCCTGCCTCCAGCTGCCCCAGCCATCTCGCAGGCGCCCGAGGCGCTGTCGCGGACCAGG 109835

Qy         979 GCCAGCACCGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGG 1038
               |||||||| ||||||||||| ||||| ||||||||||||||||| ||  |||||||||||
Db      109834 GCCAGCACTGCGCGCTTCGTTTGCCGCGCGTCCGGGGAGCCACGACCTACGCTGCACTGG 109775

Qy        1039 CTGCACGACGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGC 1098
               ||||||||||||  ||||||||||||||| ||||||||||||||||||||||||||||||
Db      109774 CTGCACGACGGGGCCCCGTTGCGACCCAACGGGCGCGTCAAGGTGCAGGGCGGTGGCGGC 109715

Qy        1099 AGCTTGGTCATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAA 1158
               ||||||||||||||||||||||||||||||||| |||| ||||||||||||||||| || 
Db      109714 AGCTTGGTCATCACTCAGATCGGCCTGCAGGACTCTGGTTACTACCAGTGCGTAGCGGAG 109655

Qy        1159 AACAGCGCGGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTG 1218
               |||||||||||||| |||||||| ||||||||||||||||| || |||||||||||||||
Db      109654 AACAGCGCGGGAACCGCCTGTGCGGCTGCGCCCCTGGCGGTGGTAGTGCGCGAGGGGCTG 109595

Qy        1219 CCCAGCGCCCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCC 1278
```

```
                 ||||| |||||||||||||||||  || |||||||||||||||||| |||||||| ||||||
Db      109594 CCCAGTGCCCCGACTCGGGTCAAGGCAACGCCGCTGAGCAGCTCCGCTGTGCTAGTGGCC 109535

Qy        1279 TGGGAGCGGCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAG 1338
               ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      109534 TGGGAGCGGCCTGAATTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAG 109475

Qy        1339 GCAAGGGGAGTGGACAATGTGGAGTACCAGTTTGCAGTAA 1378
               ||||||||  ||    | | | ||  ||  |     |||| |
Db      109474 GCAAGGGGTATGTCTACTCTTGACCACGGGACCCCAGTTA 109435


RESULT 14
AC068507/c
LOCUS       AC068507               176282 bp    DNA     linear   HTG 06-MAY-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-349H15 map 15, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC068507
VERSION     AC068507.2  GI:8389522
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349H15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176282)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 9, 2000 this sequence version replaced gi:7684421.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10275
                Center clone name: 349_H_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156826 bases at least Q40
                Consensus quality: 166608 bases at least Q30
                Consensus quality: 170674 bases at least Q20
                Insert size: 173682; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1477: contig of 1477 bp in length
            *     1478      1577: gap of 100 bp
            *     1578      2785: contig of 1208 bp in length
            *     2786      2885: gap of 100 bp
            *     2886      4859: contig of 1974 bp in length
            *     4860      4959: gap of 100 bp
            *     4960      6416: contig of 1457 bp in length
            *     6417      6516: gap of 100 bp
            *     6517      8739: contig of 2223 bp in length
            *     8740      8839: gap of 100 bp
            *     8840     10868: contig of 2029 bp in length
            *    10869     10968: gap of 100 bp
            *    10969     13737: contig of 2769 bp in length
            *    13738     13837: gap of 100 bp
            *    13838     15329: contig of 1492 bp in length
            *    15330     15429: gap of 100 bp
            *    15430     17938: contig of 2509 bp in length
            *    17939     18038: gap of 100 bp
            *    18039     19981: contig of 1943 bp in length
            *    19982     20081: gap of 100 bp
            *    20082     23587: contig of 3506 bp in length
            *    23588     23687: gap of 100 bp
            *    23688     27447: contig of 3760 bp in length
            *    27448     27547: gap of 100 bp
            *    27548     30161: contig of 2614 bp in length
            *    30162     30261: gap of 100 bp
            *    30262     34737: contig of 4476 bp in length
            *    34738     34837: gap of 100 bp
            *    34838     40101: contig of 5264 bp in length
            *    40102     40201: gap of 100 bp
            *    40202     44039: contig of 3838 bp in length
            *    44040     44139: gap of 100 bp
            *    44140     49950: contig of 5811 bp in length
            *    49951     50050: gap of 100 bp
            *    50051     55555: contig of 5505 bp in length
            *    55556     55655: gap of 100 bp
            *    55656     60496: contig of 4841 bp in length
            *    60497     60596: gap of 100 bp
            *    60597     68544: contig of 7948 bp in length
            *    68545     68644: gap of 100 bp
            *    68645     76940: contig of 8296 bp in length
            *    76941     77040: gap of 100 bp
            *    77041     85480: contig of 8440 bp in length
            *    85481     85580: gap of 100 bp
            *    85581     93905: contig of 8325 bp in length
            *    93906     94005: gap of 100 bp
            *    94006    108063: contig of 14058 bp in length
            *   108064    108163: gap of 100 bp
            *   108164    123496: contig of 15333 bp in length
            *   123497    123596: gap of 100 bp
            *   123597    139965: contig of 16369 bp in length
            *   139966    140065: gap of 100 bp
            *   140066    176282: contig of 36217 bp in length.
FEATURES             Location/Qualifiers
     source          1..176282
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349H15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1477
```

```
                    /note="assembly_fragment"
     misc_feature   1578. .2785
                    /note="assembly_fragment"
     misc_feature   2886. .4859
                    /note="assembly_fragment"
     misc_feature   4960. .6416
                    /note="assembly_fragment"
     misc_feature   6517. .8739
                    /note="assembly_fragment"
     misc_feature   8840. .10868
                    /note="assembly_fragment"
     misc_feature   10969. .13737
                    /note="assembly_fragment"
     misc_feature   13838. .15329
                    /note="assembly_fragment"
     misc_feature   15430. .17938
                    /note="assembly_fragment"
     misc_feature   18039. .19981
                    /note="assembly_fragment"
     misc_feature   20082. .23587
                    /note="assembly_fragment"
     misc_feature   23688. .27447
                    /note="assembly_fragment"
     misc_feature   27548. .30161
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:left"
     misc_feature   30262. .34737
                    /note="assembly_fragment"
     misc_feature   34838. .40101
                    /note="assembly_fragment"
     misc_feature   40202. .44039
                    /note="assembly_fragment"
     misc_feature   44140. .49950
                    /note="assembly_fragment"
     misc_feature   50051. .55555
                    /note="assembly_fragment"
     misc_feature   55656. .60496
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:right"
     misc_feature   60597. .68544
                    /note="assembly_fragment"
     misc_feature   68645. .76940
                    /note="assembly_fragment"
     misc_feature   77041. .85480
                    /note="assembly_fragment"
     misc_feature   85581. .93905
                    /note="assembly_fragment"
     misc_feature   94006. .108063
                    /note="assembly_fragment"
     misc_feature   108164. .123496
                    /note="assembly_fragment"
     misc_feature   123597. .139965
                    /note="assembly_fragment"
     misc_feature   140066. .176282
                    /note="assembly_fragment"
BASE COUNT    44272 a  43276 c  42765 g  43361 t   2608 others
ORIGIN

  Query Match             11.8%;  Score 331.2;  DB 2;  Length 176282;
  Best Local Similarity   80.6%;  Pred. No. 2.1e-58;
  Matches  387;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        867 CTGCCGAGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGT 926
Db     166357 CTGGGGGGGGGCAGAGACCTTATTTGCGGGATCCCGCGCTCAGCGTCGCCCCTCGCGCC 166298

Qy        927 GCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCAC 986
Db     166297 CCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCAC 166238

Qy        987 CGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGA 1046
Db     166237 AGCGCGCTTCGTGTGCCGCGCGTCGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAA 166178

Qy       1047 CGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGT 1106
Db     166177 CGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGT 166118

Qy       1107 CATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGC 1166
Db     166117 CATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGC 166058

Qy       1167 GGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGC 1226
Db     166057 GGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGC 165998

Qy       1227 CCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCG 1286
Db     165997 CCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCG 165938

Qy       1287 GCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGG 1346
Db     165937 GCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGG 165878


RESULT 15
AC105129
LOCUS       AC105129               198295 bp    DNA     linear   PRI 26-APR-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-349G13, complete sequence.
ACCESSION   AC105129
VERSION     AC105129.4  GI:20330981
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-349G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
```

```
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 198295)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 26, 2002 this sequence version replaced gi:20148056.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22971
                Center clone name: 349_G_13
            --------------.
FEATURES             Location/Qualifiers
     source          1. .198295
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-349G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1. .79)
                     /rpt_family="AluS"
     repeat_region   complement(183. .494)
                     /rpt_family="AluSq"
     repeat_region   complement(1216. .1416)
                     /rpt_family="L1MB8"
     repeat_region   complement(1417. .1709)
                     /rpt_family="AluSg"
     repeat_region   complement(1710. .2115)
                     /rpt_family="L1MB8"
     repeat_region   2116. .2169
                     /rpt_family="(TTTTA)n"
     repeat_region   2186. .2234
                     /rpt_family="AluSq/x"
     repeat_region   complement(2236. .2512)
                     /rpt_family="AluJb"
     repeat_region   complement(2514. .2631)
                     /rpt_family="L1MB8"
     repeat_region   complement(2632. .2941)
                     /rpt_family="AluY"
     repeat_region   complement(2942. .3104)
                     /rpt_family="L1MB8"
     repeat_region   3181. .3464
                     /rpt_family="AluSx"
     repeat_region   3465. .3711
                     /rpt_family="AluJb"
     repeat_region   3712. .3788
                     /rpt_family="(TA)n"
     repeat_region   complement(4661. .4940)
                     /rpt_family="AluJb"
     repeat_region   complement(4946. .5106)
                     /rpt_family="FRAM"
     repeat_region   complement(5878. .6193)
                     /rpt_family="AluYb8"
     repeat_region   7197. .7490
                     /rpt_family="AluSx"
     repeat_region   7925. .8226
                     /rpt_family="AluY"
     repeat_region   complement(8227. .8312)
                     /rpt_family="L1MB3"
     repeat_region   8313. .8347
                     /rpt_family="AT_rich"
     repeat_region   complement(8350. .8617)
                     /rpt_family="AluSx"
     repeat_region   complement(8621. .9274)
                     /rpt_family="L1MB3"
     repeat_region   9403. .9445
                     /rpt_family="AT_rich"
     repeat_region   9706. .10007
                     /rpt_family="AluJb"
     repeat_region   complement(10940. .11000)
                     /rpt_family="L1MB5"
     repeat_region   11001. .11284
                     /rpt_family="AluJo"
     repeat_region   complement(11285. .11366)
                     /rpt_family="L1MB5"
     repeat_region   complement(11641. .11934)
                     /rpt_family="AluSg"
     repeat_region   complement(12071. .12294)
                     /rpt_family="AluJb"
     repeat_region   complement(12410. .12697)
                     /rpt_family="AluSx"
     repeat_region   complement(12708. .12998)
                     /rpt_family="AluJb"
     repeat_region   13158. .13478
                     /rpt_family="AluJo"
     repeat_region   13970. .14093
                     /rpt_family="FLAM_C"
     repeat_region   14108. .14412
                     /rpt_family="AluJo"
     repeat_region   complement(14813. .15119)
                     /rpt_family="AluSx"
     repeat_region   complement(15130. .15432)
                     /rpt_family="AluSq"
     repeat_region   complement(15532. .15841)
                     /rpt_family="AluSx"
     repeat_region   complement(15906. .16077)
                     /rpt_family="AluSg/x"
     repeat_region   complement(16079. .16383)
                     /rpt_family="AluSq"
```

```
    repeat_region   16409. .16481
                    /rpt_family="MER5A"
    repeat_region   16482. .16779
                    /rpt_family="AluSq"
    repeat_region   16780. .16874
                    /rpt_family="MER5A"
    repeat_region   complement(17247. .17543)

  Query Match             11.8%;  Score 331.2;  DB 9;  Length 198295;
  Best Local Similarity   80.6%;  Pred. No. 2.1e-58;
  Matches  387;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        867 CTGCCGAGCCAACAAGCCCCTCACGCGTGACTTCGCCACTGCGGCTGCTGAGCTCCGAGT 926

Db      89528 CTGGGGGGGGGCAGAGACCTTAGTTGCGGGATCCCGCGCTCAGCGTCGCCCCTCGCGCC 89587

Qy        927 GCTTGCTGCCCCAGCCATCTCGCAGGCACCCGAGGCGCTCTCGCGGACGCGGGCCAGCAC 986

Db      89588 CCTAGCGGCTCCCGCCATCACTCAGGCGCCCGAGGCGCTGTCGCGGACGCGGGCGAGCAC 89647

Qy        987 CGCGCGCTTCGTGTGCCGGGCGTCCGGGGAGCCACGGCCCGCGCTGCACTGGCTGCACGA 1046

Db      89648 AGCGCGCTTCGTGTGCCGCGCGTCGGGGGAGCCGCGGCCAGCGCTGCGCTGGCTGCACAA 89707

Qy       1047 CGGGATCCCGTTGCGACCCAATGGGCGCGTCAAGGTGCAGGGCGGTGGCGGCAGCTTGGT 1106

Db      89708 CGGGGCGCCGCTGCGGCCCAACGGGCGCGTCAAGGTCCAGGGCGGCGGTGGCAGCCTGGT 89767

Qy       1107 CATCACTCAGATCGGCCTGCAGGACGCTGGCTACTACCAGTGCGTAGCAGAAAACAGCGC 1166

Db      89768 CATCACACAGATCGGCCTGCAGGACGCCGGCTACTACCAGTGCGTGGCTGAGAACAGCGC 89827

Qy       1167 GGGAACTGCCTGTGCCGCTGCGCCCCTGGCGGTAGTGGTGCGCGAGGGGCTGCCCAGCGC 1226

Db      89828 GGGAATGGCGTGCGCTGCCGCGTCGCTGGCCGTGGTGGTGCGCGAGGGGCTGCCCAGCGC 89887

Qy       1227 CCCGACTCGGGTCACAGCCACGCCGCTGAGCAGCTCCTCTGTGCTGGTGGCCTGGGAGCG 1286

Db      89888 CCCCACGCGGGTCACTGCTACGCCACTGAGCAGCTCCGCTGTGTTGGTGGCCTGGGAGCG 89947

Qy       1287 GCCTGAGTTGCACAGCGAGCAAATCATTGGCTTCTCTCTTCACTACCAAAAGGCAAGGGG 1346

Db      89948 GCCCGAGATGCACAGCGAGCAGATCATCGGCTTCTCTCTCCACTACCAGAAGGCACGGGG 90007

Search completed: October 15, 2003, 04:35:54
Job time : 10466.9 secs
```